2019

CHEMICAL SYNTHESIS, CHARACTERIZATION AND BIOLOGICAL EVALUATION OF METHYLATION AND GLYCATION DNA ADDUCTS

Qi Tang
University of Rhode Island, qitang@uri.edu

CHEMICAL SYNTHESIS, CHARACTERIZATION AND BIOLOGICAL EVALUATION OF METHYLATION AND GLYCATION DNA ADDUCTS

BY

QI TANG

A DISSERTATION SUBMITTED IN PARTIAL FULFILLMENT OF THE

REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

IN

PHARMACEUTICAL SCIENCES

UNIVERSITY OF RHODE ISLAND

2019

DOCTOR OF PHILOSOPHY DISSERTATION

OF

QI TANG

APPROVED:

Dissertation Committee:

Major Professor Deyu Li

Bongsup Cho

Gongqin Sun

Nasser H. Zawia
DEAN OF THE GRADUATE SCHOOL

UNIVERSITY OF RHODE ISLAND
2019

ABSTRACT

The integrity of genomic DNA is constantly challenged by endogenous and environmental agents with the formation of DNA adducts. Some of these adducts are toxic and mutagenic to replication, thus potentially leading to cancer and other genetic diseases. To counteract the undesired DNA modifications from damaging agents, cells have evolved a number of repair pathways, such as base-excision repair, nucleotide-excision repair, mismatch repair, and direct reversal repair, to restore the intact DNA. However, DNA repair is not always efficient, there are many interfering factors, such as inherited deficiency in DNA repair pathways, or the abnormal uptake of substance with inhibitory effects to DNA repair enzymes that may result in the accumulation of DNA adducts. In the meantime, cells have equipped with mechanisms to carry out translesion bypass of DNA adducts, and the bypass of these adducts may lead to mutagenesis. Therefore, studying the DNA adduct formation, repair, and other mutagenic consequences can shed light on how DNA damaging agents impact on cellular response of organisms. A comprehensive understanding of the biological outcomes of interested DNA adducts requires the development of a set of efficient chemical approaches to prepare and characterize adduct-containing DNA oligonucleotides. On the other hand, the biological evaluation of these DNA adducts from various aspects, such as studying their genotoxic effects, and exploring potential interfering factors to their cellular repair, is essential for providing insights into the etiology of many diseases including cancer.

This dissertation describes the chemical synthesis, characterization, and biological evaluation of methylation and glycation DNA adducts by using a variety of

chemical and genetic tools. These strategies along with the findings obtained from the application of them are briefly discussed in the abstract of four manuscripts as following, and are described in detail in **CHAPTER 1, 2, 3, and 4**.

In **MANUSCRIPT-I**, the objective of this study was to develop a rigorous procedure to chemically synthesize and characterize adduct-containing DNA oligonucleotides for biological studies. Oligonucleotides serve as important tools for biological, chemical, and clinical research. The preparation of oligonucleotides through automated solid-phase synthesis is well-established. However, identification of byproducts generated from DNA synthesis, especially from oligonucleotides containing site-specific modifications, is sometimes challenging. Typical high-performance liquid chromatography, mass spectrometry, and gel electrophoresis methods alone are not sufficient for characterizing unexpected byproducts, especially for those having identical or very similar molecular weight to the products. We developed a rigorous quality control procedure to characterize byproducts generated during oligonucleotide syntheses: (1) purify oligonucleotides by different HPLC systems; (2) determine exact molecular weight by high-resolution MS; (3) locate modification position by MS/MS or exonuclease digestion with matrix-assisted laser desorption ionization-time of flight analysis; and (4) conduct, where applicable, enzymatic assays. We applied these steps to characterize byproducts in the syntheses of oligonucleotides containing biologically important methyl DNA adducts 1-methyladenine and 3-methylcytosine. In 1-methyladenine synthesis, we differentiated a regioisomeric byproduct 6-methyladenine, which possesses a molecular weight identical to uncharged 1-methyladenine. As for 3-methylcytosine, we identified a

deamination byproduct 3-methyluracil, which is only 1 Da greater than uncharged 3-methylcytosine in the ~4900 Da context. The detection of these byproducts would be very challenging if the abovementioned procedure was not adopted.

In **MANUSCRIPT-II**, as we have developed a platform enabling us to conduct *in vitro* synthesis and quality control of interested DNA adducts, we synthesized a number of N-methyl DNA adducts occurring at the Watson-Crick base-pairing face of the four nucleobases including 1-methyladenine, 3-methylcytosine, 1-methylguanine, and 3-methylthymine. We evaluated the repair preference of AlkB family DNA repair enzymes on these N-methyl DNA adducts under difference strand context. The AlkB protein is a repair enzyme that uses an α-ketoglutarate/Fe(II)-dependent mechanism to repair alkyl DNA adducts. AlkB has been reported to repair highly susceptible substrates, such as 1-methyladenine and 3-methylcytosine, more efficiently in single strand-DNA than in double strand-DNA. Here, we tested the repair of weaker AlkB substrates 1-methylguanine and 3-methylthymine and found that AlkB prefers to repair them in double strand-DNA. We also discovered that AlkB and its human homologues, ALKBH2 and ALKBH3, are able to repair the aforementioned adducts when the adduct is present in a mismatched base pair. These observations demonstrate the strong adaptability of AlkB toward repairing various adducts in different environments.

In **MANUSCRIPT-III**, we aimed to explore the potential interfering effects of the overconsumption of substance from environmental resources to DNA repair. We studied the inhibitory activity of a class of natural products, hydrolysable tannins, on ALKBH2 enzyme. Hydrolysable tannins are a class of polyphenolic compounds

commonly found in many plants. In this work, we studied the *in vitro* inhibitory mechanism of six molecules on ALKBH2. We determined the IC50 values of these compounds on the repair of 3-methylcytosine, the proto-typical substrate of ALKBH2. A structure-activity relationship was also observed between the strength of inhibition and the number of galloyl moieties in a molecule. In addition, we found that the inhibition by this class of polyphenolic compounds on ALKBH2 is through an iron-chelating mechanism. Direct reversal of alkyl DNA adducts by AlkB family enzymes constitute an important cellular repair mechanism for maintaining genome integrity. We here demonstrated that the potential effect of overdosing hydrolysable tannins might lead to the inhibition of iron-dependent AlkB family enzymes.

In **MANUSCRIPT-IV**, the objective of this work was to explore the *in vitro* synthesis and the in cell biological outcomes of glycation DNA adducts generated from cellular reducing sugar glucose-6-phosphate. We chemically synthesized and characterized the Amadori glycation products from glucose-6-phosphate and evaluated their toxic and mutagenic properties in cell. Reducing sugars and their metabolic derivatives (e.g. D-glucose, D- glucose-6-phosphate, methylglyoxal, and glyoxal) can non-enzymatically react with cellular biomacromolecules, such as proteins and nucleic acids, to induce glycation. Glucose and glucose-6-phosphate have been shown to glycate lysine residues of protein to form Amadori adducts. As a glycating agent with multiple-fold higher reactivity and much more abundant cellular concentration than glucose, glucose-6-phosphate has also been proposed to glycate DNA and cause genotoxicity with the formation of Amadori adducts. Due to the unstable nature of Amadori DNA adducts and a number of potential repair pathways, it is challenging to

study their toxic and mutagenic properties in cell. Therefore, the specific biological outcomes of these adducts remain unclear. In this work, we used a number of chemical and genetic approaches and studied the biological consequences, such as replication block and mutations, of Amadori DNA adducts in a site-specific manner. Our data showed that Amadori DNA adducts arising from glucose-6-phosphate can induce G to T and single nucleotide deletion, which may lead to devastating biological consequences if left unrepaired, thereby affecting the genome integrity of organisms.

In summary, this dissertation offers a systematic strategy to conduct both *in vitro* and in cell studies of DNA adducts and mutagenesis. The methods applied in this work will not only broaden the basic research for chemical modification of nucleic acids, but also bring new aspects for studying DNA biology.

ACKNOWLEDGMENTS

I would like to thank my Ph.D. supervisor, Dr. Deyu Li, for his expertise, guidance, and assistance throughout my graduate career at University of Rhode Island. The professionalism and enthusiasm he passed onto me, always encouraged me to become an independent scientist equipped with curiosity, passion, and vision. I am deeply grateful for his kindness and patience teaching me all the valuable skills in exploring unknowns, writing proposals, carrying out research, and most importantly, being an ambitious researcher with the goal of contributing knowledge to improve people's health benefits. Without his help, my achievements would not have been possible.

I would like to extend my sincere gratitude to my doctoral committee members, Dr. Bongsup Cho, Dr. Ying Zhang, and Dr. Gongqin Sun. Whenever I need help, they are always there for me. I truly appreciate their support, insights, and advice for my research projects, and serving on my committee.

I want to express my heartiest acknowledgements to Dr. Xiaoqun Dong and Dr. Qiushi Lin for their guidance and inspiration in my first year of graduate study at URI.

The accomplishment of this dissertation was possible because of the help I received from my cooperative and resourceful labmates. Their efforts and inputs enabled me to complete this work. I am particularly thankful to Dr. Fangyi Chen, Ke Bian for their invaluable support over the years.

In addition, I thank Dr. Al Bach and Ms. Kim Andrew at Rhode Island INBRE Laboratory for their help in training me on various research instruments and giving me

advice when I was having a hard time. Their help in every aspect of my research was imperative to my completion of the Ph.D. degree

Besides, the assistance and cooperation of my friends, faculty, staff at URI and collaborative institutes were essential for my completion of this work. Dr. Hang Ma, Dr. Ang Cai, Dr. Yongqiang Liu, Dr. Zhengxi Wei, and Dr. Matthew Bertin provided me with tremendous help in my research related to analytical chemistry and molecular biology. Ms. Patricia Ryan Murray and Ms. Gerralyn Perry offered their kindest efforts on helping me solving problems from lab purchasing and financial administration. Ms. Kathleen Hayes has been assisting my graduate study since my first day of being accepted as a graduate student at College of Pharmacy. I apologize for not being able to list more names, but I remember all the help they offered me.

This dissertation is dedicated to my parents, Mrs. Aixin Zhou and Mr. Huiyan Tang, and to my family for their endless love, support, and encouragement.

PREFACE

This dissertation was written according to the University of Rhode Island "Guidelines for the Format of Theses and Dissertations" standards for manuscript format. This dissertation comprises of four manuscripts that have been combined to meet the degree requirements of the Department of Biomedical and Pharmaceutical Sciences, College of Pharmacy, University of Rhode Island.

Two manuscripts have been published under the following titles:

I. Characterization of byproducts from chemical syntheses of oligonucleotides containing 1-methyladenine and 3-methylcytosine.

   Published on *ACS OMEGA* 2017, 2, 8205-8212

II. Adaptive response enzyme AlkB preferentially repairs 1-methylguanine and 3-methylthymine adducts in double-stranded DNA.

   Published on *Chemical Research in Toxicology* 2016, 29, 687-693

Two additional manuscripts will be submitted under the following titles:

III. Hydrolysable tannins are iron chelators that inhibit AlkB family DNA repair enzyme ALKBH2.

   Submitted on *Chemical Research in Toxicology*

IV. Guanine DNA glycation by glucose 6-phosphate induces G to T mutation and single nucleotide deletion in cell

TABLE OF CONTENTS

LIST OF TABLES

MANUSCRIPT-I

MANUSCRIPT-II

MANUSCRIPT-III

MANUSCRIPT-IV

N/A

LIST OF FIGURES

MANUSCRIPT-I

MANUSCRIPT-II

MANUSCRIPT-IV

CHAPTER 1

CHARACTERIZATION OF BYPRODUCTS FROM CHEMICAL SYNTHESES OF OLIGONUCLEOTIDES CONTAINING 1-METHYLADENINE AND 3-METHYLCYTOSINE

[PUBLISHED ON ACS OMEGA 2017, 2, 8205-8212]

BY

Qi Tang,[†,‖] Ang Cai,[†,‖] Ke Bian,[†] Fangyi Chen,[†,§] James C. Delaney, [‡] Sravani Adusumalli,[†] Alvin C. Bach, II,[†] Fatemeh Akhlaghi,[†] Bongsup P. Cho,[†] and Deyu Li[*,†]

[†]Department of Biomedical and Pharmaceutical Sciences, College of Pharmacy, University of Rhode Island, Kingston, Rhode Island 02881, United States.

[‡]Visterra Inc., One Kendall Square, Cambridge, Massachusetts 02139, United States

Present Address:

[§]Xiamen University, 4221 Xiang An South Road, Xiang An District, Xiamen, Fujian 361102, P.R. China (F.C.).

[‖]Q.T. and A.C. contributed equally to this work.

[*]Corresponding Author

E-mail: deyuli@uri.edu

ABSTRACT

Oligonucleotides serve as important tools for biological, chemical, and medical research. The preparation of oligonucleotides through automated solid-phase synthesis is well established. However, identification of byproducts generated from DNA synthesis, especially from oligonucleotides containing site-specific modifications, is sometimes challenging. Typical HPLC, Mass Spectrometry (MS), and gel electrophoresis methods alone are not sufficient for characterizing unexpected byproducts, especially for those having identical or very similar molecular weight (MW) to the products. We used a rigorous quality control procedure to characterize byproducts generated during oligonucleotide syntheses: 1) purify oligonucleotide by different HPLC systems; 2) determine the exact MW by high resolution MS; 3) locate modification position by MS/MS or exonuclease digestion with MALDI-TOF analysis; and 4) conduct, where applicable, enzymatic assays. We applied these steps to characterize byproducts in the syntheses of oligonucleotides containing important methyl DNA adducts 1-methyladenine (m1A) and 3-methylcytosine (m3C). In m1A synthesis, we differentiated a regioisomeric byproduct 6-methyladenine, which possesses identical MW to m1A. As for m3C, we identified a deamination byproduct 3-methyluracil, which is only 1 Dalton greater than m3C in the ~ 4,900 Dalton context. The detection of these byproducts would be very challenging if the abovementioned procedure were not adopted.

INTRODUCTION

Oligonucleotides synthesized chemically are widely used as drugs and research tools in biology, chemistry, and medicine. Solid phase synthesis of DNA and RNA

oligonucleotides has been well developed and automated phosphoramidite-based chemical process has become highly efficient.[1–3] Phosphoramidites of regular and some modified nucleotides are commercially available and certain oligonucleotide strands could be readily obtained from commercial sources.[4–7] Besides the great development in the synthesis of oligonucleotides, the differentiation of byproducts from the product oligonucleotides, especially on those containing site-specifically modified structures, is sometimes ignored by the end users. If those byproducts or small impurities were not identified and removed, it could have devastating consequences for the corresponding biological assays and medical treatments.[8,9] For certain instances, it is challenging to identify some byproducts generated in the synthesis and deprotection steps, especially those byproducts that have identical or very similar MW to the desired product oligonucleotide.

We have synthesized various oligonucleotides containing modified structures in the past, focusing on alkyl or aryl DNA adducts, by using solid and solution phase DNA synthesis.[10–14] In the syntheses of oligonucleotides containing 1-methyladenine (m1A) and 3-methylcytosine (m3C), we observed byproducts are either regioisomer (identical MW) to the product or have 1 Dalton in MW greater than the product in a ~ 4,900 Dalton context (Figure 1a). m1A and m3C are formed by exogenous and endogenous alkylating agents mainly in single-stranded DNA and have been detected both *in vitro*[15–21] and *in vivo*[15,20,22–29]. Both adducts are cytotoxic and block DNA replication, and are the best substrates for the AlkB family DNA repair enzymes (Figure 1b).[30–34] m3C has also been proposed an epigenetic biomarker for cancer.[35]

Most of our DNA syntheses successfully provided target oligonucleotides. However, some oligonucleotides contained side reaction contaminants generated during the synthesis or deprotection steps, requiring further purification and identification. To that end, we applied a rigorous quality control process which entails the following steps: i) Synthesize an adduct-containing oligonucleotide from the corresponding phosphoramidite; ii) purify the product oligonucleotide by both reverse phase and anion exchange HPLC; iii) measure the exact MW of the oligonucleotide by high resolution MS (HR-MS), certain impurities can also be detected by HR-MS; iv) determine the modification position by either MS/MS or exonuclease digestion with MALDI-TOF analysis; and v) test the biological activity of the adduct-containing oligonucleotide by appropriate enzymatic assays. We are assuming that the modified nucleotide has been extensively investigated for its stability under solid-phase chemical synthesis and deprotection beforehand; however, the strategy outlined above should confirm this. Nuclear Magnetic Resonance (NMR; ^{1}H, ^{13}C, and ^{31}P) is a powerful tool for structural integrity studies of the phosphoramidite on the milligram scale.[36–38] However, NMR becomes impractical for characterizing minor impurities within the modified structures in synthetic oligonucleotides due to increasing spectral complexity, as well as the final amount of material available (e.g. on a micro-mole or less scale). Below, we report the detailed characterization of byproducts in the syntheses of oligonucleotides containing m1A and m3C. The characterization and separation of those byproducts provides confidence in the quality of oligonucleotides used in further biological experiments.

EXPERIMENTAL SECTION

Synthesis of Oligonucleotides Containing m1A and m3C. Oligonucleotides (16mer) with the sequence 5'-GAAGACCTXGGCGTCC-3' containing the lesions at X position were made by using solid-phase phosphoramidite chemistry on a MerMade-4 Oligonucleotide Synthesizer.[10,34,39] The phosphoramidites were purchased from ChemGenes. N1-Methyl deoxyadenosine (n-fmoc) phosphoramidite was used for m1A synthesis. N3-Methyl deoxycytidine (n-bz) phosphoramidite was used for m3C synthesis. The final cleavage and deprotection step was usually carried out by treating the oligonucleotide with concentrated aqueous ammonium hydroxide (28%) at 80 ℃ for 3 hours. The modified cleavage and deprotection step for minimizing the byproduct formation was at 25℃ for 16 h. The concentration of an oligonucleotide was determined by measuring UV absorbance at 260 nm. The extinction coefficient (ε) of a certain adduct is calculated as its unmodified counterpart due to the slight variation between the values in the context of a 16mer DNA.

Purity Test of Oligonucleotides Containing m1A and m3C and Related Byproducts by HPLC. The purity of oligonucleotides was tested by both reverse phase (C18) and anion exchange HPLC methods. For the reverse phase test, liquid chromatographic separation was performed by using a Phenomenex Luna Semi-Preparative C18 column (9 x 250 mm, 5 μm) at a flow rate of 3 mL/min. Solvent A was 100 mM triethylammonium acetate in water and solvent B was 100% acetonitrile. A solvent gradient was carried out under the following conditions: 2.0% of B for 0.2 min, 2.0 to 9.0% of B over 0.3 min, 9.0 to 9.4% of B over 16 min, 9.4 to 70.0% of B over 0.5 min, 70.0% of B for 3 min, 70.0 to 2.0% of B over 0.5 min, and 2.0 % B over 4.5 min. The column oven was set at 40 ℃. The UV signal at 260 nm was used to

detect the oligonucleotide absorbance. For the anion exchange LC analysis, oligonucleotides were purified and tested by using a Thermo DNAPac PA-100 anion exchange column (4 x 250 mm, 13 µm) with solvent A as water and solvent B as 1.5 M ammonium acetate in water. A solvent gradient was carried out under the following conditions: 50.0% of B for 1 min, 50.0 to 52.0% of B over 2 min, 52.0 to 75.0% over 1 min, 75.0% of B for 2 min, 75.0 to 50.0% of B over 0.5 min, and 50.0% of B for 4.5 min. The flow rate was at 4.0 ml/min.

HPLC-MS Analysis. High resolution MS analyses of oligonucleotides were performed on AB Sciex triple quadrupole-TOF mass spectrometers (ABI4600 and ABI5600). ESI was conducted by using a needle voltage of 4.0 kV in a negative ion mode. A heated capillary was set at 600 ℃. The nebulizer gas pressure was 40 psi; the heater gas pressure was 40 psi; the curtain gas pressure was 25 psi; the declustering potential was -220 V; and the collision energy was -10 V. Liquid chromatographic separation was performed using a Phenomenex Luna C 18 column (4.6 × 100 mm; 5 µm) at a flow rate 0.4 mL/min. Solvent A was 10 mM ammonium acetate in water and solvent B was 100% acetonitrile. A solvent gradient was carried out under the following conditions: 2.0% of B for 0.5 min, 2.0 to 17.4% of B over 11min, 17.4 to 60.0% of B over 0.1 min, 60.0% of B for 2 min, 60.0 to 2.0% of B over 0.1 min, and 2.0 % B over 3.3 min.

MS/MS Analysis. Oligonucleotide fragmentation analyses were performed by manually injecting the oligonucleotide samples of 100 pmol/µl into AB Sciex triple quadrupole-TOF mass spectrometers (ABI4600 and ABI5600). The syringe flow rate was set at 10 µl/min. ESI was conducted by using a needle voltage of 4.5 kV in a

negative ion mode. A heated capillary was set at 400 ℃. The nebulizer gas pressure was 75 psi; the heater gas pressure was 25 psi; the curtain gas pressure was 25 psi; the declustering potential was -100 V; and the collision energy was -15 V. The parent ion *m/z* for m6A, m1A, m3C and m3U were selected as 816.14 (-6 charge), 816.14 (-6 charge), 974.77 (-5 charge) and 974.96 (-5 charge), respectively. Data analyses were performed with the AB Sciex Analyst TF software 1.7.

Exonuclease Digestion with MALDI-TOF Analysis. The modified oligonucleotides were characterized by 3'-5'exonuclease digestion followed by MALDI-TOF analysis. In general, 1.0 μL of sample containing 200-250 pmol of a modified oligonucleotide was used for digestion. SVP (0.2 unit) was added together with 6.0 μL of ammonium citrate (100 mM, pH 9.4) and 6.0 μL of water for the 3′ to 5′ digestion.[11] For the MALDI-TOF analyses, 1.0 μL of the mixture was used at a certain time point until the digestion was almost finished; the digestion was quenched by mixing with 1.0 μL of MALDI analysis matrix (3-hydroxypicolinic acid and di-ammonium hydrogen citrate in a 1:1 ratio). Samples were analyzed by a Shimadzu Axima Performance MALDI-TOF mass spectrometer using a 50Hz laser with a power setting of 90 in positive ion reflection mode with 500 shots collected.[13]

Enzymatic Reaction with the AlkB protein and HPLC Analysis. All reactions were performed at 37 ℃ for 1 h in a reaction buffer containing 70.0 μM $Fe(NH_4)_2(SO_4)_2 \cdot 6H_2O$, 0.93 mM α-ketoglutarate, 1.86 mM ascorbic acid, and 46.5 mM HEPES (pH 8.0).[10] The reactions were quenched by adding 10.0 mM EDTA followed by heating at 95 ℃ for 5 min. Typically, 0.25 μM of the purified AlkB protein was incubated with 5.0 μM m6A and m1A in the presence of all cofactors in a

20 μL reaction volume. For the m3C and m3U, 0.18 μM of AlkB was incubated with 5.0 μM of oligonucleotides. Samples were then analyzed under an HPLC condition that was able to separate the substrate and product.

To separate the starting material and product of the enzymatic reactions for the four lesions, different HPLC conditions were optimized by using either C18 or anion exchange columns. The UV detection was set at 260 nm. Specifically, (1) m6A and its repaired product A were analyzed by using a Phenomenex Luna Semi-Preparative C18 column (10 x 150 mm, 5 μm) under the following conditions: Solvent A was 100 mM triethylammonium acetate in water and solvent B was 100% acetonitrile. A solvent gradient was carried out under the following conditions: 2.0% of B for 0.2 min, 2.0 to 9.0% of B over 0.3 min, 9.0 to 9.4% of B over 16 min, 9.4 to 70.0% of B over 0.5 min, 70.0% of B for 3 min, 70.0 to 2.0% of B over 0.5 min, and 2.0 % B for 4.5 min. The flow rate was at 3.0 ml/min. The column oven was set at 40 ℃. This protocol was also used to analyze m3U and its repair product U. (2) m1A and its repaired product A were analyzed by using a Thermo DNAPac PA-100 anion exchange column (4 x 250 mm, 13 μm) with solvent A as water and solvent B as 1.5 M ammonium acetate in water. A solvent gradient was carried out under the following conditions: 50.0% of B for 1 min, 50.0 to 52.0% of B over 2 min, 52.0 to 75.0% over 1 min, 75.0% of B for 2 min, 75.0 to 50.0% of B over 0.5 min, and 50.0% of B for 4.5 min. The flow rate was at 4.0 ml/min. (3) m3C and its repaired product C were analyzed with a similar protocol to m1A/A analysis, except the solvent gradient was changed to the following conditions: 50.0% of B for 1 min, 50.0 to 57.0% of B

over 1 min, 57.0% of B for 2 min, 57.0 to 50.0% of B for 0.5 min, then 50.0% of B for 4.5 min.

RESULTS

Identification of m6A Byproduct from m1A Synthesis. We were trying to incorporate m1A into an oligonucleotide as a substrate for the AlkB repair reaction.[34,40] A 16mer oligonucleotide containing m1A was prepared by using the commercially available phosphoramidite of m1A (Figure 1a). The final product was deprotected under standard conditions by treating the crude oligonucleotide with ammonium hydroxide at 80 ℃ for 3h. The oligonucleotide was then tested by both reverse phase and anion exchange HPLC (see Experimental Procedures for detailed conditions). The resulting chromatograms showed a single peak under both conditions (retention time 10.2 min in Figure 2a and 4.0 min in Figure S22). High resolution ESI-TOF MS analysis of the sample exhibited *m/z* at 1224.715 at its -4 charge state, which is in good agreement with the theoretical *m/z* 1224.711 expected of the product oligonucleotide (Figure 3a, S1 and Table 1).

Collision-induced-dissociation (CID) MS/MS analysis was used to determine the location of the alkyl adduct in the synthesis of m1A-containing 16mer oligonucleotide (Figure 4). The MS/MS results, presented in detail in the supporting information (Figures S5 and Tables S1), confirmed that modification occurred at the eighth position from the 3' end of the oligonucleotide and the modified base had the same MW as m1A. Analysis of 3' end fragmentations indicated that w_{15} to w_8 fragments all contain an extra methyl modification (Figure 4, Table S1 and Figure S10). On the

other hand, w_7 to w_1 ions showed only standard DNA sequences without methyl modification (Table S1). These results suggested the modification is located between w_8 and w_7, which is the eighth position from the 3' end of the 16mer oligonucleotide. From the 5' fragmentations, we observed the methyl modification in ions from (a_{15}-C) to (a_{10}-G), but no such modification from ion (a_9-X) to (a_2-A) (Figure 4, Figure S9 and Table S1). These results suggested the modification is located between (a_{10}-G) and (a_9-X), which is the ninth position from the 5' end (Figure 4). The fragmentation patterns from both the 5' and 3' ends are consistent with the proposed m1A-containing DNA sequence (Figure 4).

To further confirm the location of the modified base, we performed enzyme-digestion coupled with MALDI-TOF MS analysis.[11–13,41,42] The procedure has been widely used to confirm the location of lesion positions by determining the mass changes after partial digestion from the 3' end by the exonuclease SVP.[43–45] Figure 5 shows the positive ion MALDI-TOF MS spectra of the 3' to 5' SVP exonuclease digestion of the product oligonucleotide at different time intervals. The *m/z* of 4904.2 at 0 min (before digestion, as a control) indicated the 16mer DNA sequence contained a methylated DNA base (theoretical *m/z* = 4902.9 Dalton, herein theoretical *m/z* values are shown in parentheses after the observed *m/z* values). After 1 minute SVP digestion, while the product ion disappeared, three new lower masses appeared at *m/z* = 4615.7 (4614.8), 4326.7 (4325.8), and 4022.7 (4021.7), which correspond to the 15, 14, and 13mer fragments generated from 3' cleavage, respectively (Figure 5). The signal intensity of three new peaks were significantly increased with the disappearance of the original product oligonucleotide ion at *m/z* = 4904.2.

Further digestion resulted in a number of smaller fragments. In 3 min digestion (Figure 5 and S17), the signal of *m/z* = 3075.1 (3074.6) matches the sequence for the fragment that was cleaved one base before the lesion. The *m/z* = 2746.4 (2745.5, containing a methyl modification), represents the digestion occurring on the 3' side of the modified adenine position (Figure 5), which also persisted in the 5 min digestion spectrum. The signal at 2419.1 (2418.5, no methyl modification) in the 3 and 5 min digestions represents the fragment generated after liberation of the deoxymethyladenosine 5' monophosphate. Digestions practically stopped at *m/z* = 1223.6/8 (1222.3), which represent the *m/z* of a 4mer nucleotide of the 5' end. Taken together, these results confirm that the modified structure is a methylated adenine and it is at the ninth position of the proposed oligonucleotide from the 5'to 3' direction.

The chromatographic and mass spectral evidence presented above indicates that the product oligonucleotide could have the MW expected for a 16mer containing the m1A adduct in the correct position. We then tested the repair of this product by the AlkB protein, which is an enzyme that repairs various alkylated DNA adducts (Figure 1b).[32,34] The result showed the repair efficiency of this oligonucleotide by AlkB was very low, even when 2.5 uM of AlkB was incubated with 5.0 uM product oligonucleotide (Figure 6b and Figure S24). This observation was in contrast to the report that m1A is a good substrate of AlkB (eg. 0.2 μM of AlkB is able to repair 5.0 μM m1A completely in 1 h).[10] We suspected that the structure in the product oligonucleotide may not be m1A, even though it had an identical MW to m1A. The reverse phase HPLC results also showed an anomalous phenomenon. For example, the positively charged m1A nucleobase[34] should be more polar than the natural base

adenine and thus have a shorter retention time on a C18 column. However, the opposite was observed: adenine has ashorter retention time than the product (Figure 2a and 2c). It has been reported that 6-methyladenine (m6A) could be generated from a Dimroth rearrangement of m1A under basic conditions(Figure 1a and Table 1).[34,46]

We therefore conducted a new batch synthesis of m1A under a milder deprotection condition (25 ℃ for 16 h). The product oligonucleotide was eluted as a single peak in both HPLC columns (Figure 2b for C18 and Figure 6d for anion exchange). The HPLC retention time in reverse phase HPLC exhibited a shorter retention time (8.5 min, Figure 2b) than the non-modified control adenine (9.0 min, Figure 2c) or the previous product (10.2 min, m6A, Figure 2a), in good agreement with the theory that m1A should be more polar than both A and m6A.

The MW of the new product was characterized by HR-MS and had the correct *m/z* of m1A at 1224.715 (1224.711) (Figure 3b, S2 and Table 1). The MS/MS results (e.g., the w_7, w_8, a_9-X, and a_{10}-G ions, Figure 4, Figure S6, S11, S12, and Table S2) demonstrated the modification is at the ninth position from the 5' end and it is a methylated adenine. The exonuclease digestion with MALDI-TOF analyses (Figure S18 and S21a) also confirmed the location and identity of adduct. For example, the 2746.1 (before adduct) and 2418.8 (after adduct) peaks at 3 min digestion in Figure S18 were present. Finally, incubation with AlkB showed an excellent repair efficiency against the new product oligonucleotide: 5.0 μM adduct was mostly repaired in 1 hr with 0.2 μM enzyme (Figure 6e).[10] Taken together, these results confirmed the newly synthesized product as a 16mer oligonucleotide containing m1A at the ninth position from the 5' end (Figure 1a).

Identification of m3U Byproduct from m3C Synthesis. We also wanted to synthesize an oligonucleotide containing m3C as the substrate for the AlkB family DNA repair reactions within the same sequence context as for the m1A oligonucleotide (Figure 1a).[34,40] After automated synthesis and standard deprotection (80 ℃ for 3h), the HPLC chromatogram from a C18 column showed two peaks with similar intensities (9.5 min and 11.6 min, Figure 2d), which were readily separated (Figure 2e and 2f). HR-MS analysis showed that the early eluting peak (9.5 min, Figure 2e) had an *m/z* value 1218.703 at -4 charge state (Figure 3c and S3), which is in agreement with that of the *m/z* of the target m3C containing oligonucleotide (theoretical *m/z* 1218.709, Table 1). The second species at 11.6 min (Figure 2f) showed an *m/z* 1218.957 at its -4 charge state (Figure 3d and S4). Compared to the 9.5 min peak, the *m/z* value of the second species was 0.254 unit higher at the -4 charge state, indicating that the MW (4,879.850 Dalton) of the second species is 1.016 Dalton higher than the first one. This is consistent with the *m/z* value of an oligonucleotide containing m3U at its -4 charge state (1218.955, Table 1). m3U could potentially be generated from deamination of m3C under basic conditions (Figure 1a).[47] The MS/MS results of the two oligonucleotides (e.g., the w_7, w_8, a_9-X, and a_{10}-G ions, Figure 4, Figure S7, S8, S13 to S16, Table S3, and S4) demonstrated the modifications are at the ninth position from the 5' end. The exonuclease digestion with MALDI-TOF analyses (Figure S19, S20 and S21b) also confirmed the location and identity of the adducts. At 7 minute digestion, signature peaks were 2722.2 (before adduct) and 2418.8 (after adduct) for m3C (Figure S18); and 2723.3 (before adduct) and 2418.7 (after adduct) for m3U (Figure S19). The C18 HPLC retention

time of the two species showed m3C containing oligonucleotide had a shorter retention time (9.5 min, Figure 2e) than m3U (11.6 min, Figure 2f). This result indicated m3C is more polar than m3U, which supports that m3C is positively charged at pH 7.0 (Figure 1a).[34] The AlkB reaction (0.2 μM) on the m3C oligonucleotide showed a good repair efficiency: 5.0 μM m3C adduct was mostly repaired to cytosine after 1 h (Figure 6h). This was not the case for m3U containing oligonucleotides under a similar condition, which showed almost no repair to m3U (Figure 6k). These observations were consistent with the observations that m3C is a strong substrate but m3U is a weak substrate for AlkB under both *in vitro* and *in vivo* conditions.[34,48]

To avoid the formation of m3U, we carried out a second batch synthesis of m3C containing oligonucleotide under a milder deprotection condition (25 ℃ for 16 h). This time we observed only the m3C containing species in the HPLC chromatogram (similar to Figure 2e). The identity of the newly synthesized m3C was confirmed by the aforementioned spectroscopic and enzymatic tests. These results indicate that the formation of the byproduct m3U was from deamination of m3C catalyzed by harsh deprotection conditions (high temperature and alkaline).[47,48] Serendipitously, the formation of m3U in the first batch of synthesis provides a new method for chemically preparing m3U-containing oligonucleotide.

DISCUSSION

For the synthesis of m1A, we observed a byproduct m6A, which exhibited identical MW and location to m1A. It is difficult to differentiate m1A and m6A without conducting bioassays and careful HPLC retention analysis. As for m3C, it

was the m3U byproduct that showed a 1 Dalton difference in the ~ 4,900 Dalton context (Table 1). The two oligonucleotides could be eluted as a single species in HPLC under certain conditions, thus it is imperative to perform high resolution MS analysis and enzymatic reaction to distinguish the two products. For HPLC analysis of modified oligonucleotides, we used a combination of two different separation systems, such as reverse phase (e.g. C18) and anion exchange chromatography. For example, it is not easy to find a suitable condition for fully separating 16mer containing m1A and A with C18 columns (Figure 2b and 2c). However, it is relatively easy to separate them with an anion exchange column, since m1A is positively charged at neutral pH but A is not (Figure 1a, 6d and 6f).[34] On the other hand, it is not easy to distinguish m6A and A under anion exchange condition (Figure S22 and S23) because of their neutrality at pH 7.0, but it is possible to separate them on the C18 column (Figure 2a and 2c).

The MW of the oligonucleotide can be determined by HR-MS, and the location of the modification can be identified by MS/MS or exonuclease digestion with MALDI-TOF analysis. To further confirm the identity of the oligonucleotide from a biological activity perspective, we recommend running enzymatic reaction on the product. In this work, we used SVP exonuclease to digest all oligonucleotides formed in the syntheses. A specific enzyme for certain types of modification could greatly help elucidate the structure of modification. So we tested the repair efficiency of m1A and m3C and their byproducts by the AlkB repair enzyme.[32,34] The AlkB protein in *E. coli* is an α-ketoglutarate/Fe(II)-dependent dioxygenase that repairs various alkyl DNA adducts, including m1A and m3C (Figure 1).[30–32,40,49–51] We found that m1A and m3C

are good substrates but m6A and m3U are weak substrates of AlkB.[34,48,49] For other modified structures, it is highly recommended that a biological or enzymatic assay should be adopted for identifying the product. The reason for adopting an enzymatic test is because a byproduct (e.g. m6A) may have an identical MW as the target product (e.g. m1A), which may be hard to differentiate by LC and MS analyses, including the MS/MS and exonuclease digestion with MALDI-TOF analysis.

CONCLUSION

Chemical synthesis of oligonucleotides is important for conducting various biological, chemical, and medical research including oligonucleotide drug development. The field of oligonucleotide synthesis has progressed such that DNA and RNA containing standard bases can be ordered from commercial sources. Most of these oligonucleotides are in high quality, but it is a good practice to perform quality control in order to confirm the products have the correct sequences in high purity. A simple HPLC/MS analysis should be sufficient for most applications. However, a more rigorous and stringent quality control procedure[52–54] should be adopted for site-specifically modified oligonucleotides, such as epigenetic marks, DNA adducts, and drug candidates. In some cases, byproducts are generated from side chemical reactions during standard automated synthesis and deprotection steps. For this reason, specific deprotection conditions may be required for preparation of oligonucleotide containing modifications, such as epigenetic biomarkers: 5-hydroxymethyl-dC, 5-formyl-dC, and 5-carboxy-dC.[55,56] After synthesis and deprotection, it is even more important to carry out a thorough purification and characterization procedure to ensure 1) complete removal of protecting groups, 2) product having high purity, and 3)

modification having the correct position and identity. In this study, we used a reliable and robust procedure to characterize byproducts from m1A and m3C syntheses. The protocol used here could be helpful for identifying byproducts generated from other oligonucleotide syntheses.

FIGURES

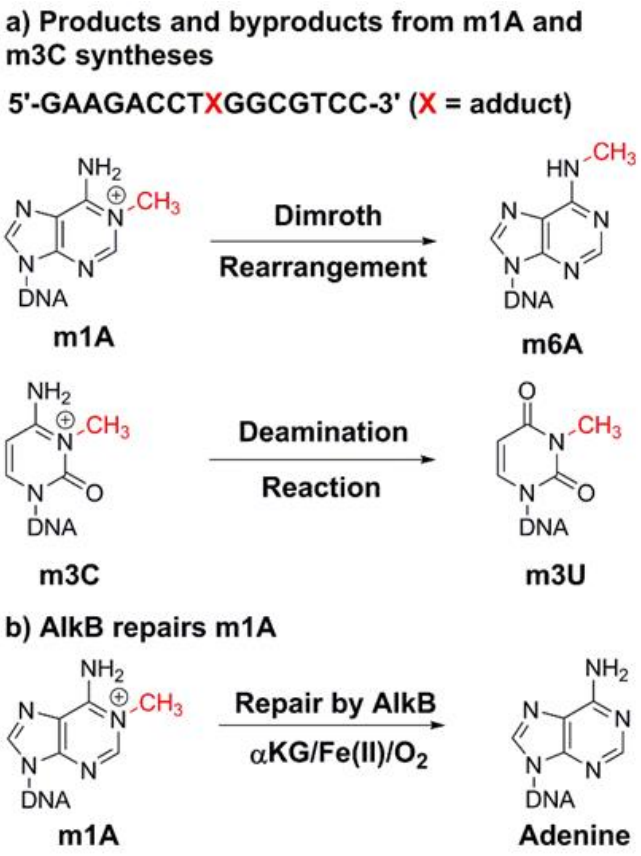

Figure 1. Oligonucleotide products and byproducts studied in this work. a) the structures of adducts and byproducts from chemical syntheses and b) the alkyl adduct m1A, as an example, is repaired in the presence of the AlkB enzyme and necessary cofactors.

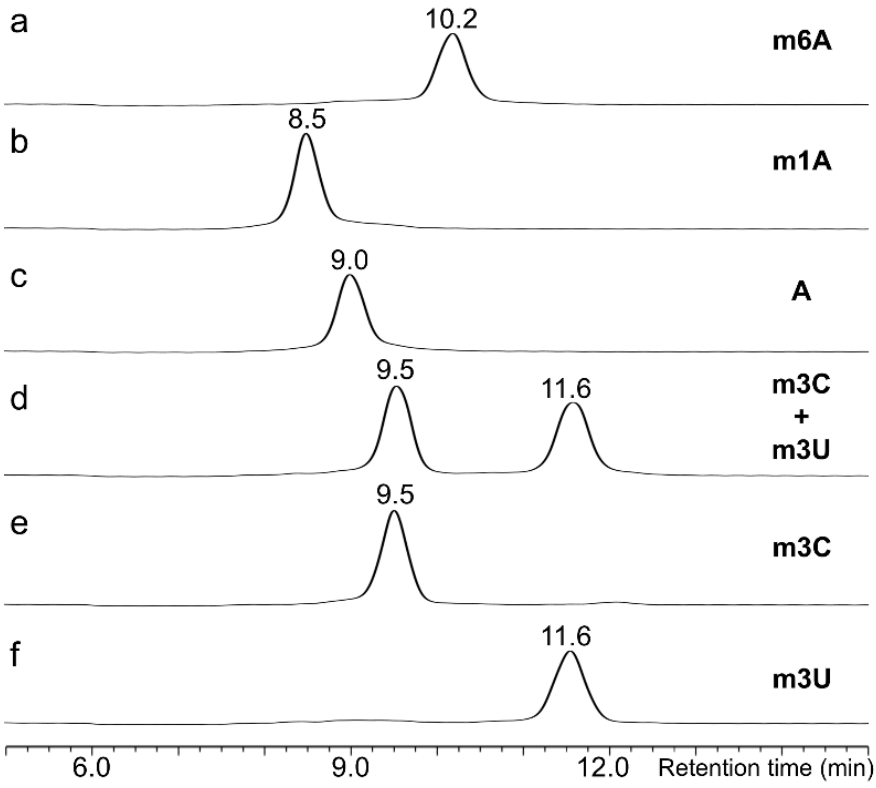

Figure 2. Reverse phase HPLC chromatograms of 16mer DNA alkyl products and byproducts. The retention time of a modification-containing oligonucleotide is labeled on top of the corresponding chromatogram. The modifications are: a) m6A); b) m1A; c) A; d) mixture of m3C + m3U; e) m3C; and f) m3U.

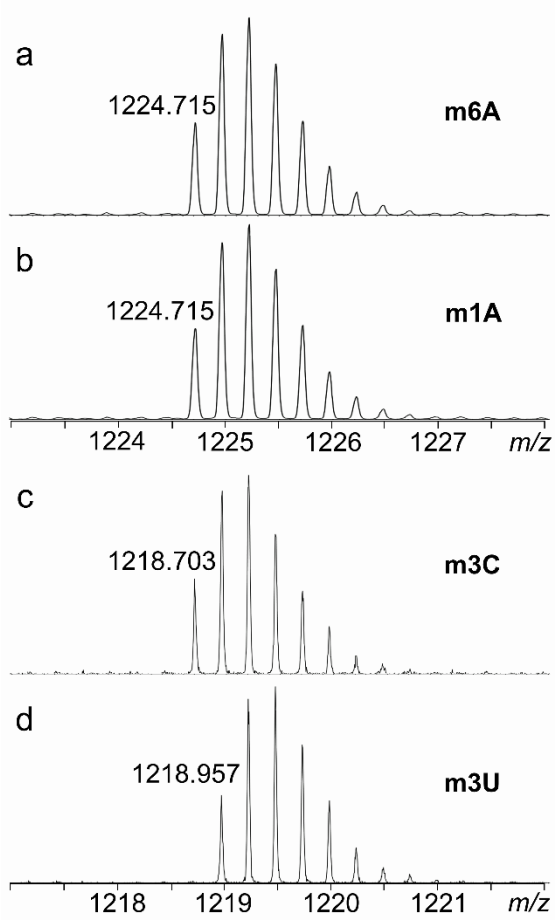

Figure 3. High resolution ESI-TOF MS analyses of 16mer DNA oligonucleotides containing target modifications and byproducts. Data represent the -4 charge envelopes and the monoisotopic peak (all 12C, 14N, etc.) values are labeled above the first peak in each peak envelope. a) The oligonucleotide (containing m6A) generated from initial m1A synthesis; b) m1A; c) m3C; and d) m3U.

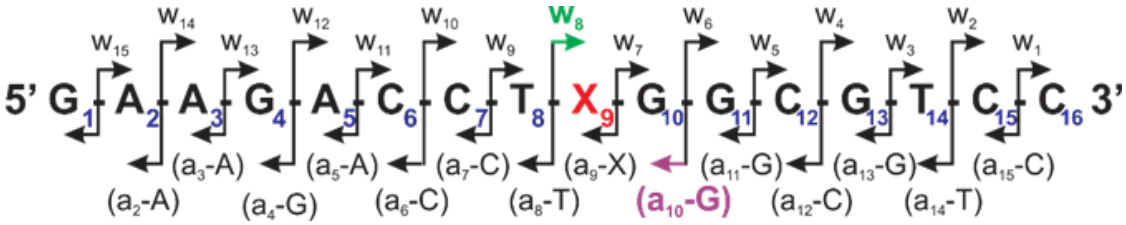

Figure 4. Predicted fragmentation pattern from collision-induced-dissociation of the 16mer oligonucleotide products from the m1A and m3C syntheses. X denotes the modified nucleotides.

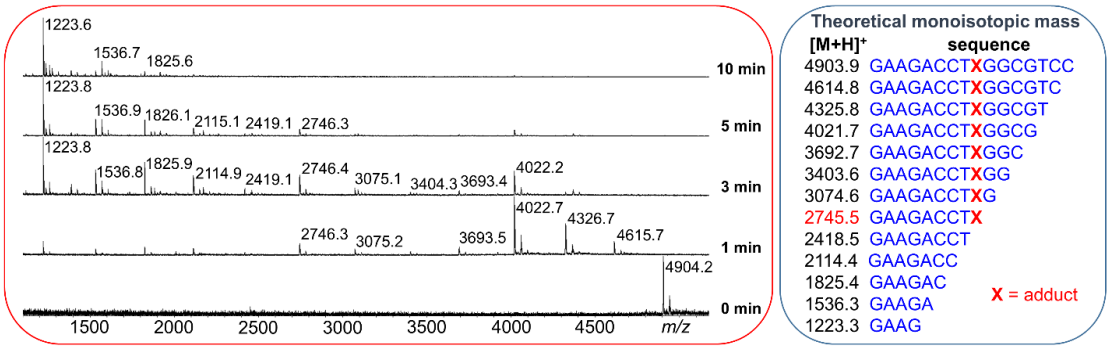

Figure 5. Time-course MALDI-TOF analyses of SVP digestion products of the 16mer product oligonucleotide (containing m6A) generated from m1A synthesis. Theoretical masses are listed in the inset. The theoretical monoisotopic mass at 2745.5 is highlighted in red because it is the smallest digestion product containing the modification.

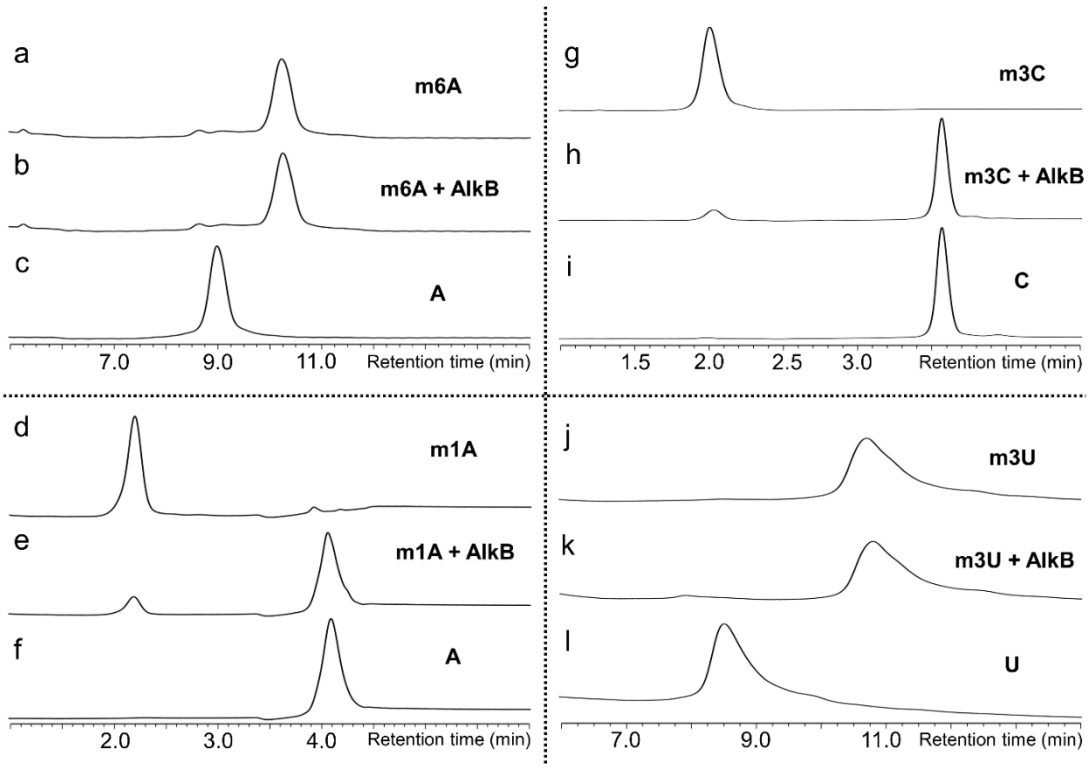

Figure 6. HPLC profiles of the AlkB repair reactions on different alkyl substrates. The three chromatograms within each panel represent one set of repair reaction including the oligonucleotides of starting material, reaction mixture, and product. For example, in the panel containing chromatograms a, b, & c, a) represents the starting material m6A, b) represents the repair reaction of m6A by AlkB, and c) represents the pure product adenine synthesized separately. Chromatograms shown in a, b, & c relevant to m6A repair and in j, k, & l relevant to m3U repair were analyzed under reverse phase conditions. Chromatograms shown in d, e, & f relevant to m1A repair and in g, h, & i relevant to m3C repair were analyzed under anion exchange conditions (see Experimental Procedures for detailed information). a) m6A; b) m6A+AlkB; c) A; d) m1A; e) m1A+AlkB; f) A; g) m3C; h) m3C+AlkB; i) C; j) m3U; k) m3U+AlkB; and l) U.

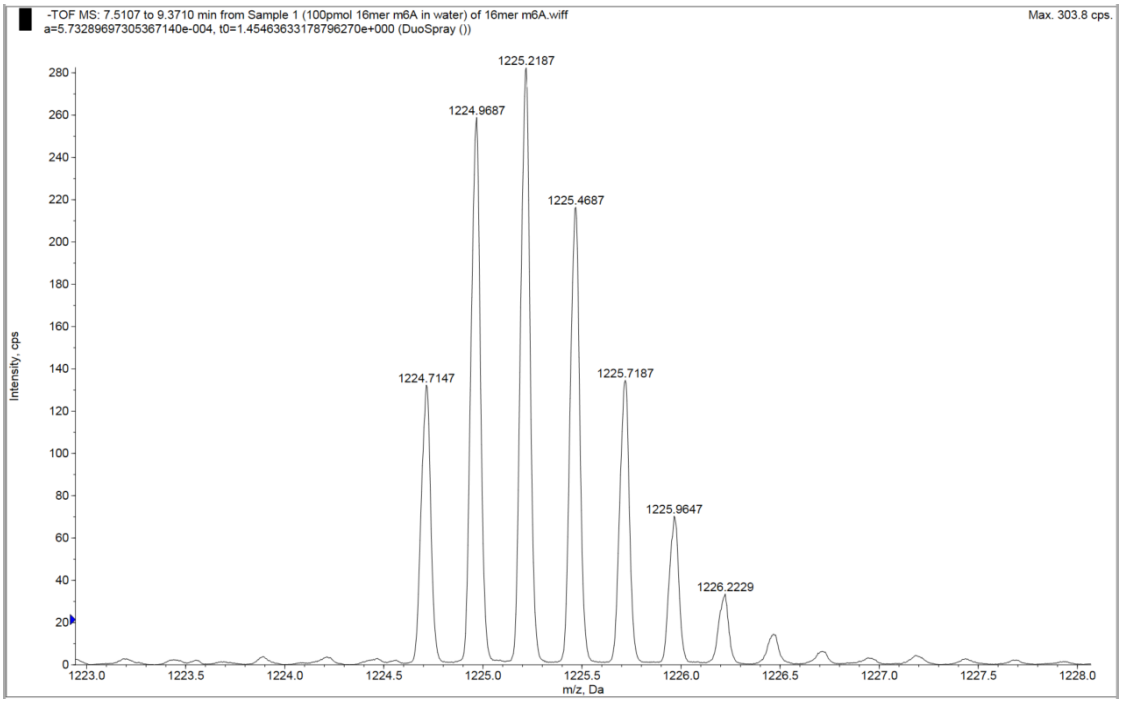

Figure S1. ESI-TOF analysis of 16mer oligonucleotide containing m6A at -4 charge state.

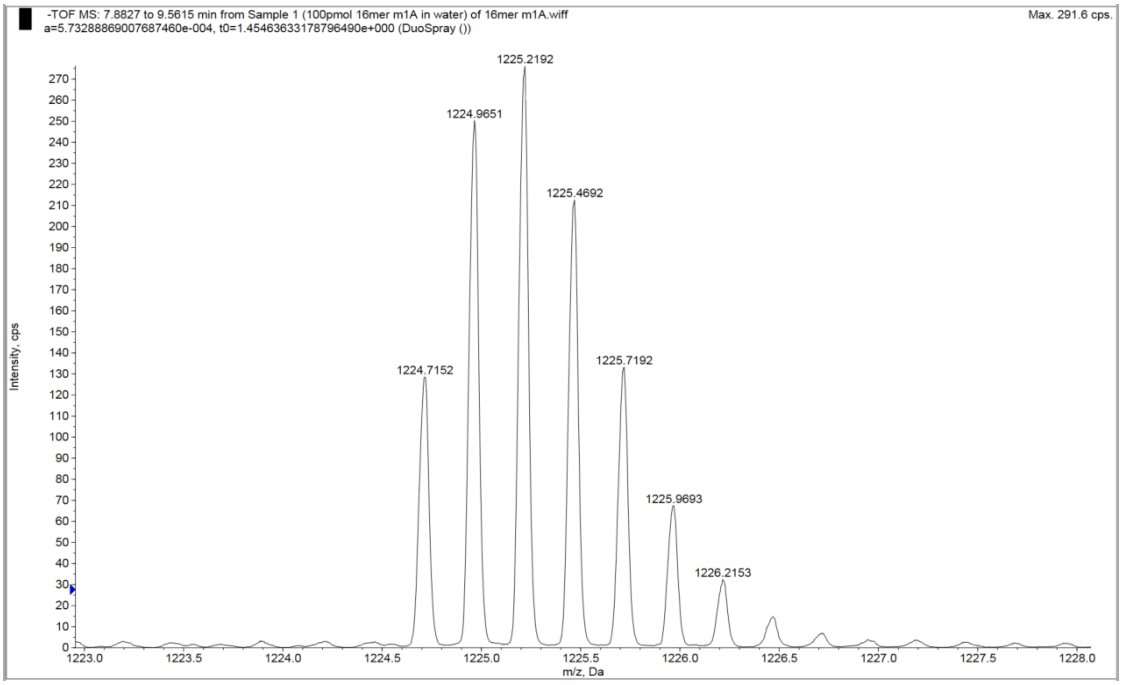

Figure S2. ESI-TOF analysis of 16mer oligonucleotide containing m1A at -4 charge state.

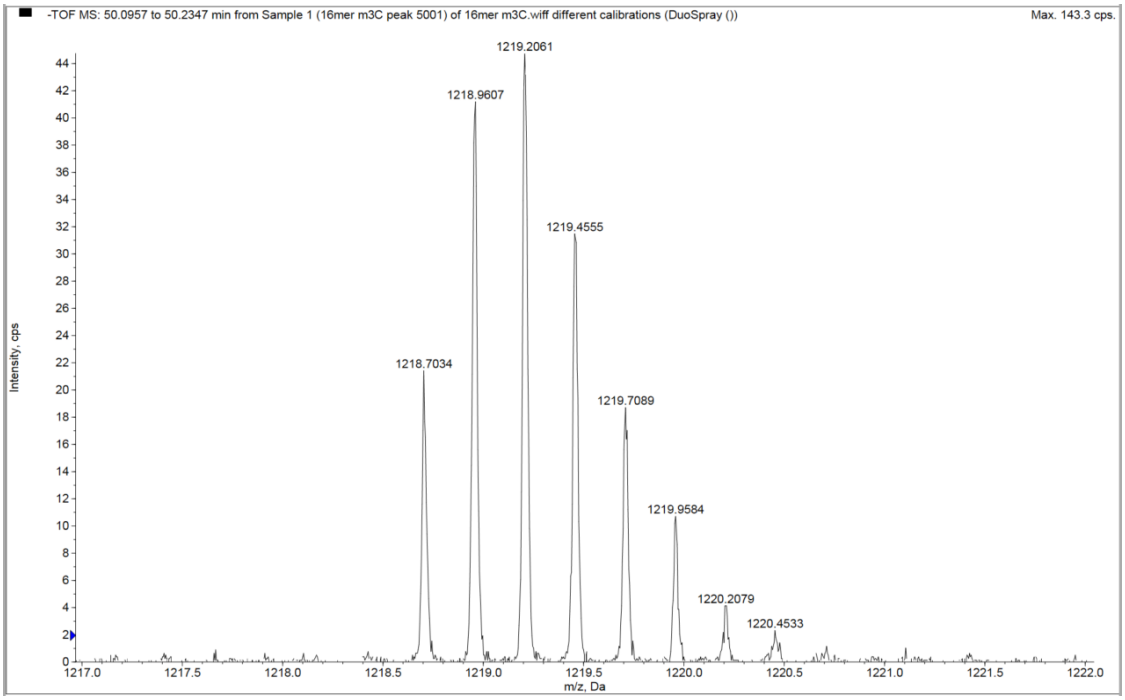

Figure S3. ESI-TOF analysis of 16mer oligonucleotide containing m3C at -4 charge state.

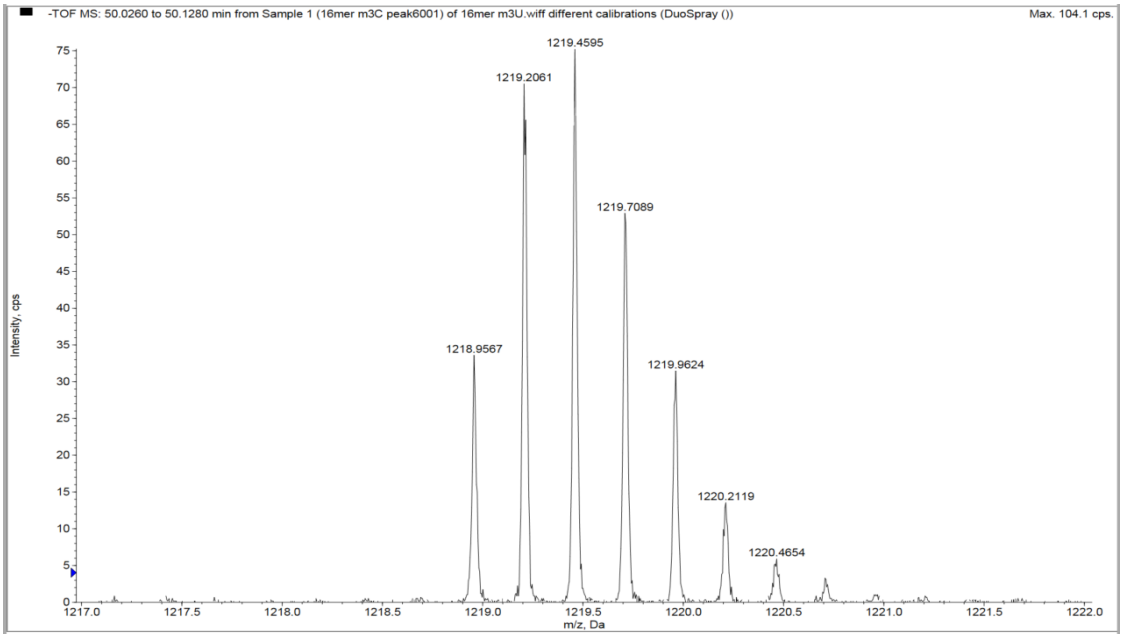

Figure S4. ESI-TOF analysis of 16mer oligonucleotide containing m3U at -4 charge state.

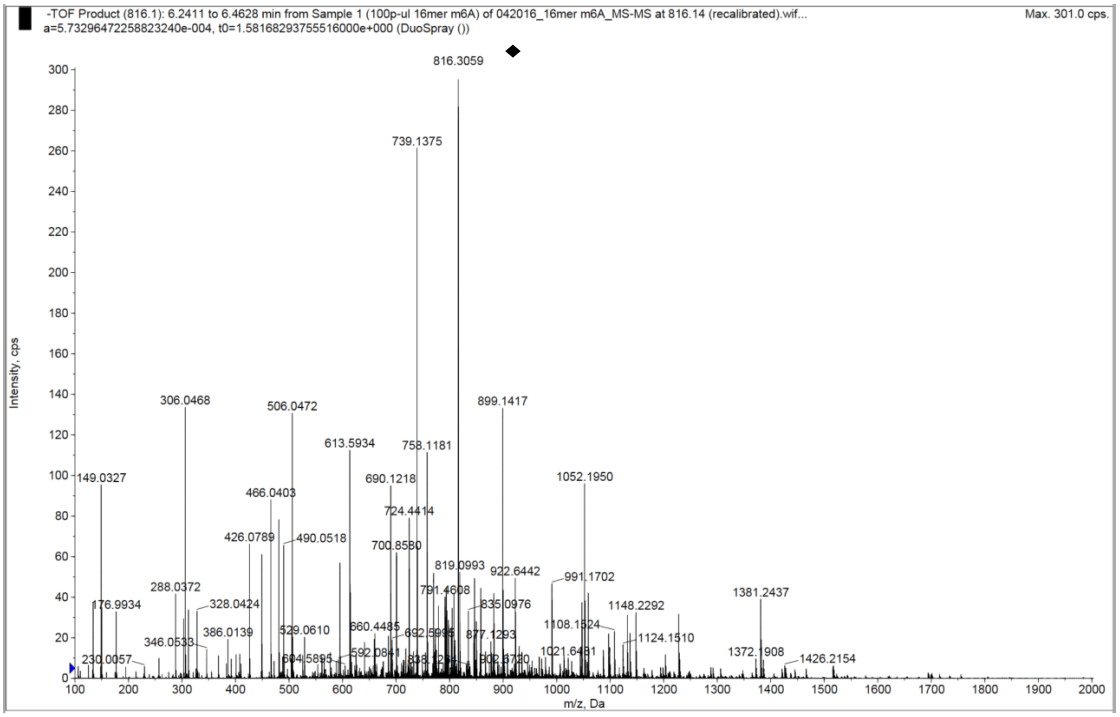

Figure S5. MS/MS fragmentation analysis of 16mer containing m6A. The diamond marks the parent ion.

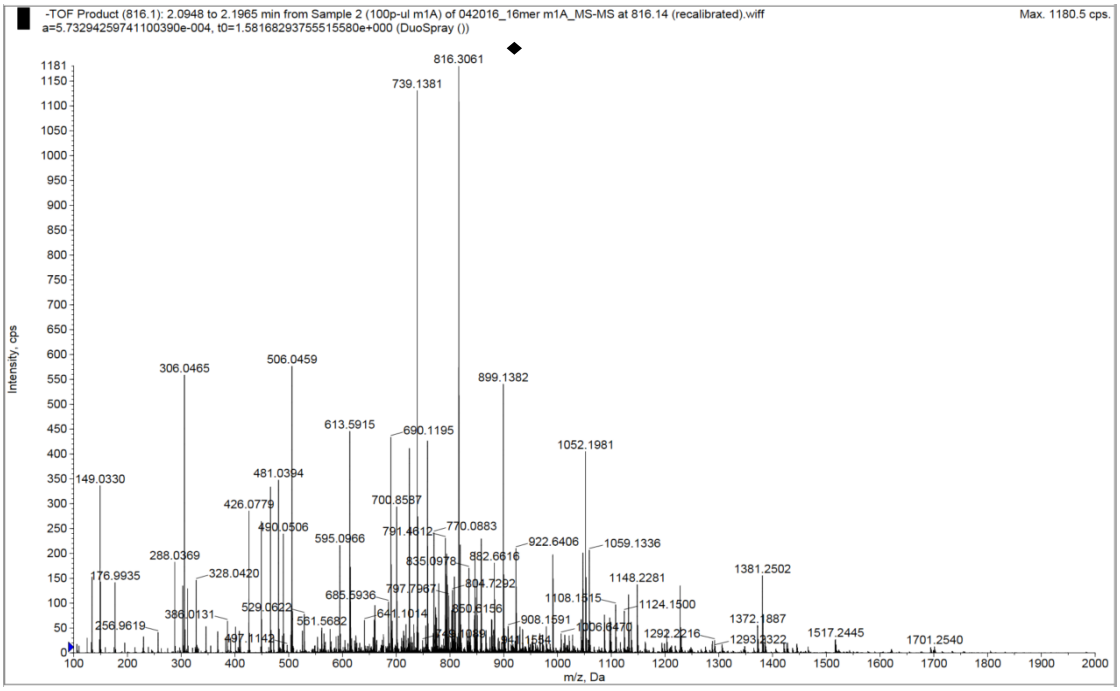

Figure S6. MS/MS fragmentation analysis of 16mer containing m1A. The diamond marks the parent ion.

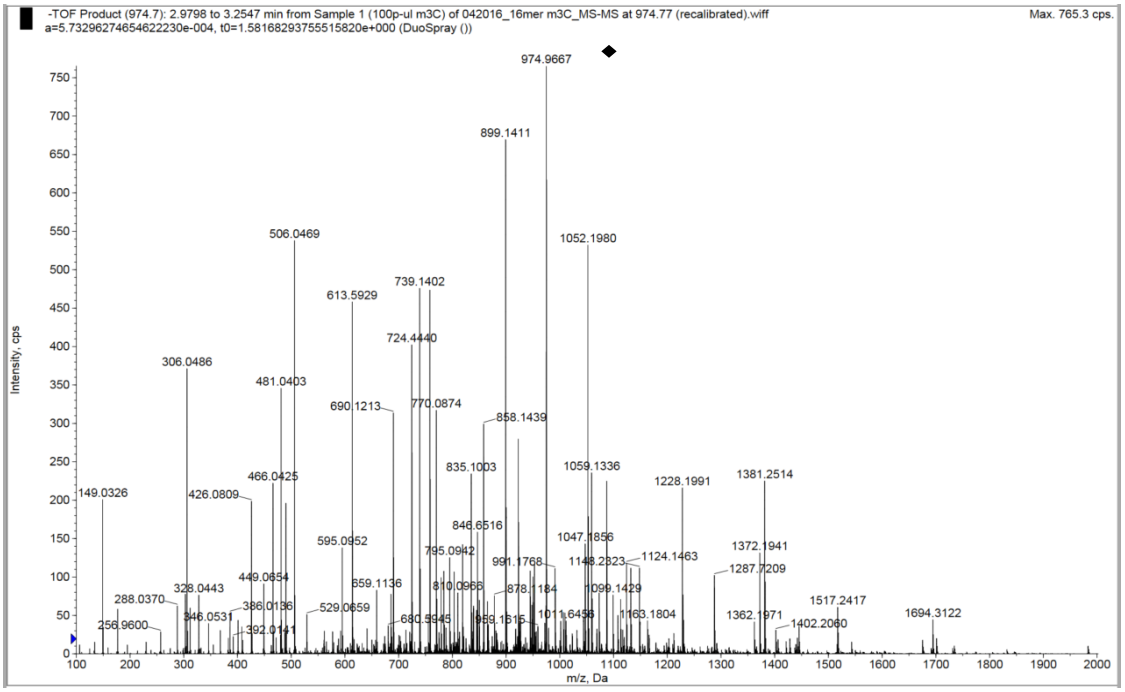

Figure S7. MS/MS fragmentation analysis of 16mer containing m3C. The diamond marks the parent ion.

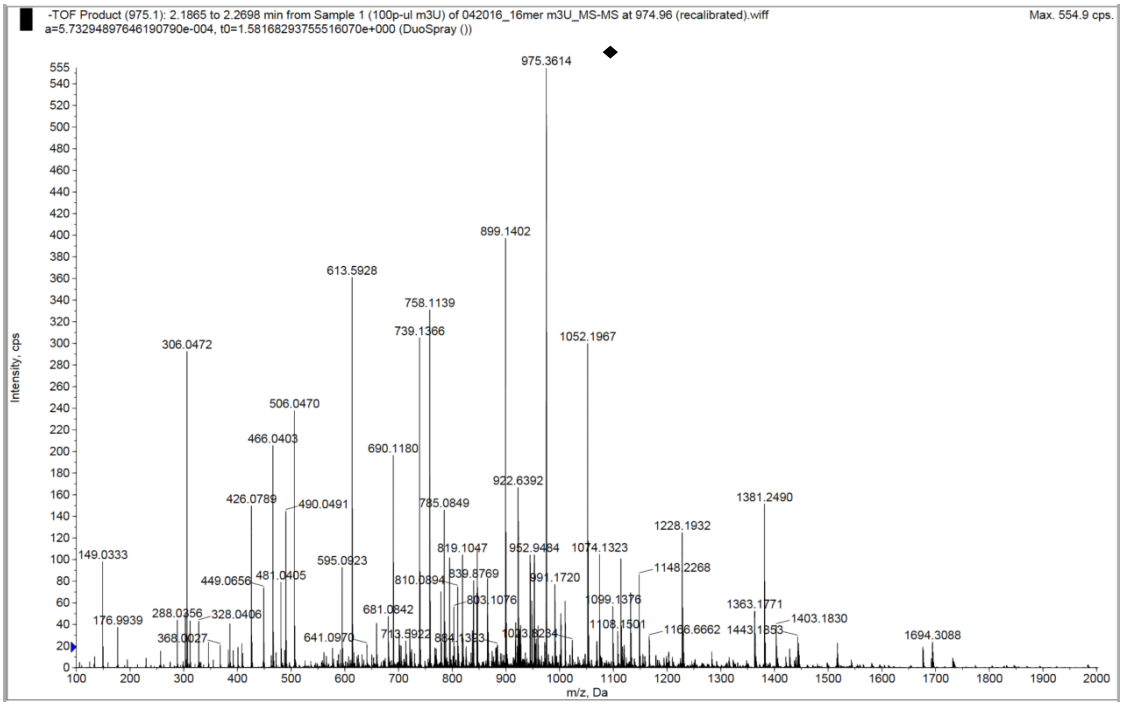

Figure S8. MS/MS fragmentation analysis of 16mer containing m3U. The diamond marks the parent ion.

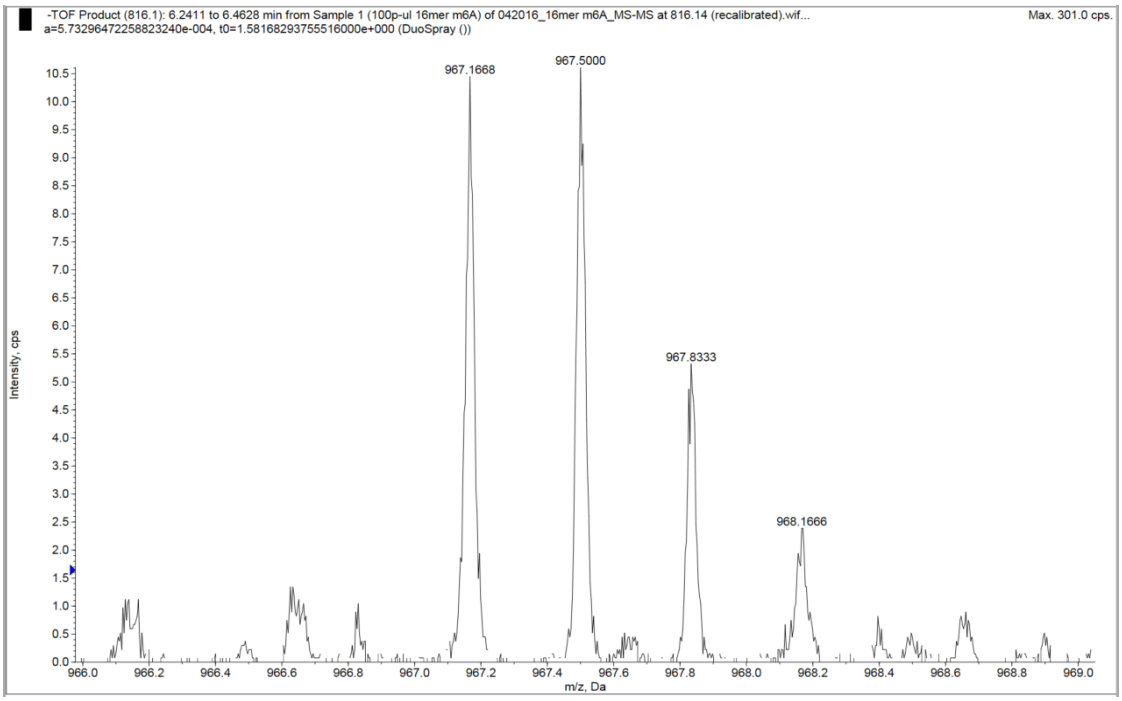

Figure S9. MS/MS fragmentation spectrum of 16mer containing m6A. The peak envelope indicates the a10-G ion at -3 charge state.

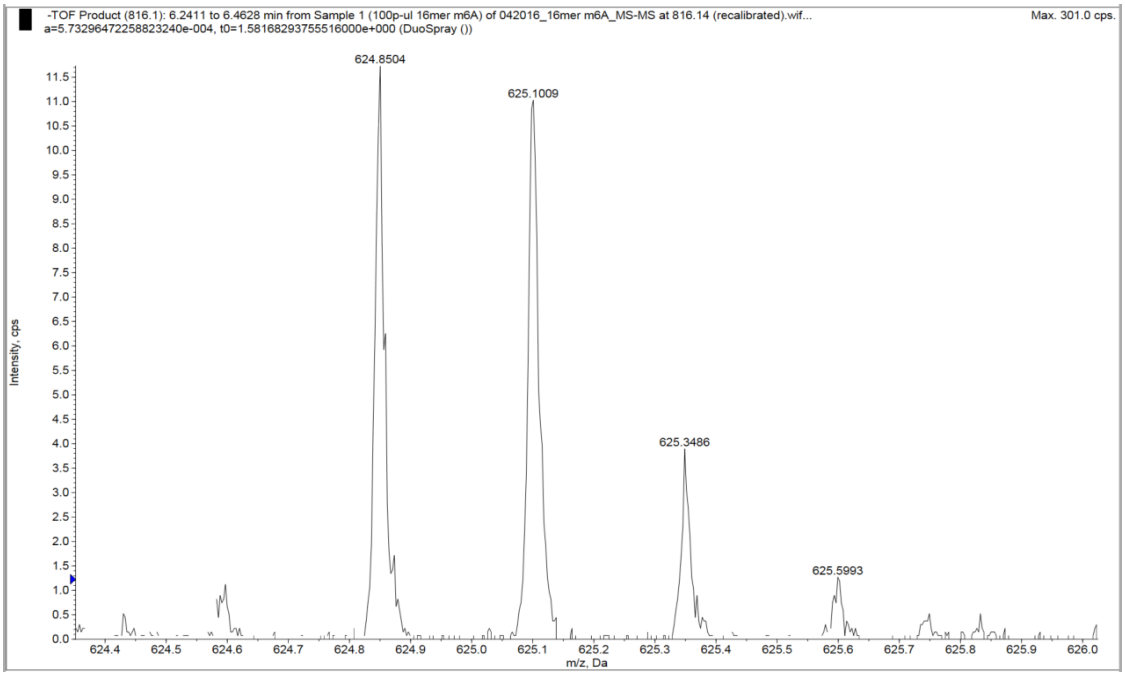

Figure S10. MS/MS fragmentation spectrum of 16mer containing m6A. The peak envelope indicates the W8 ion at -4 charge state.

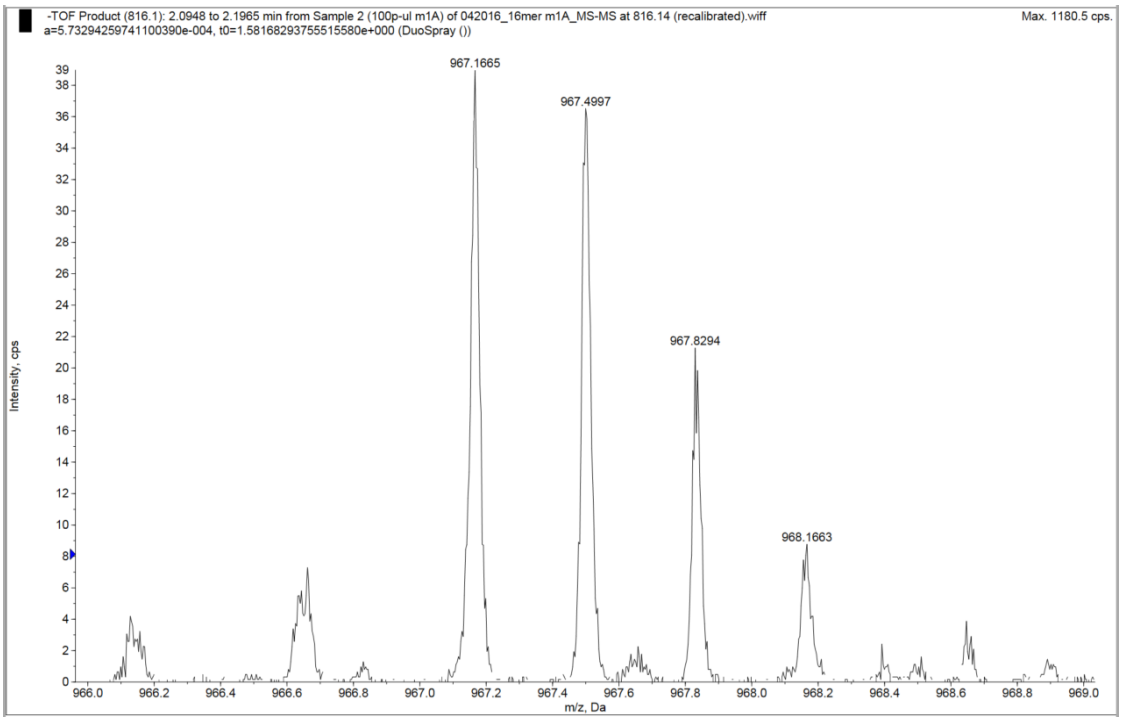

Figure S11. MS/MS fragmentation spectrum of 16mer containing m1A. The peak envelope indicates the a10-G ion at -3 charge state.

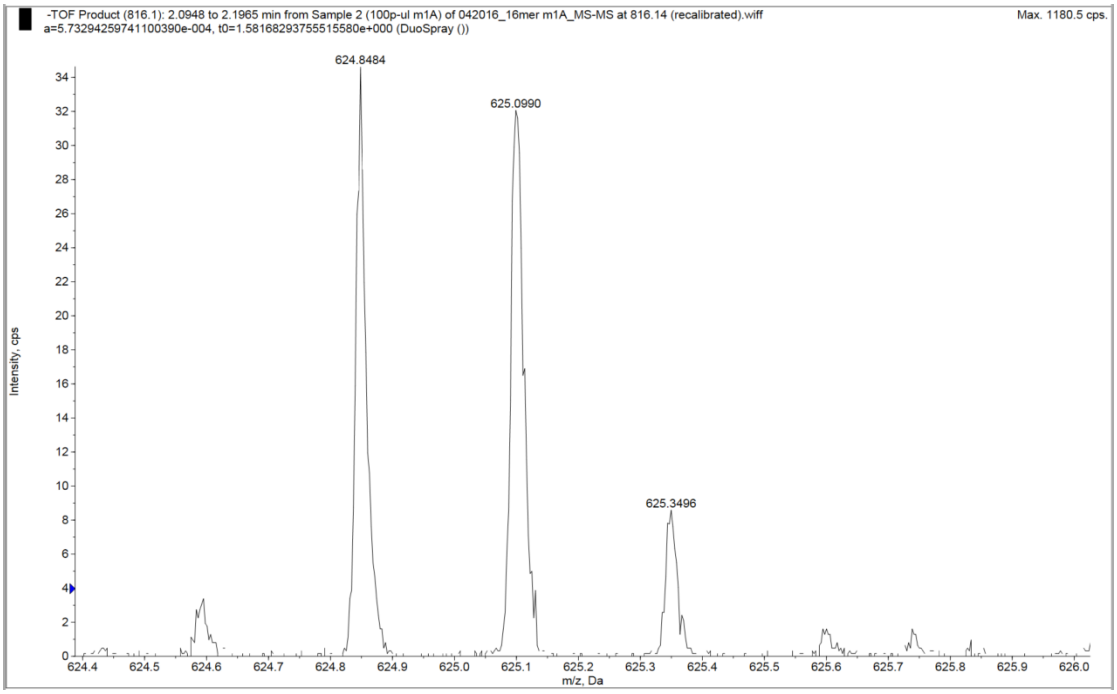

Figure S12. MS/MS fragmentation spectrum of 16mer containing m1A. The peak envelope indicates the W8 ion at -4 charge state.

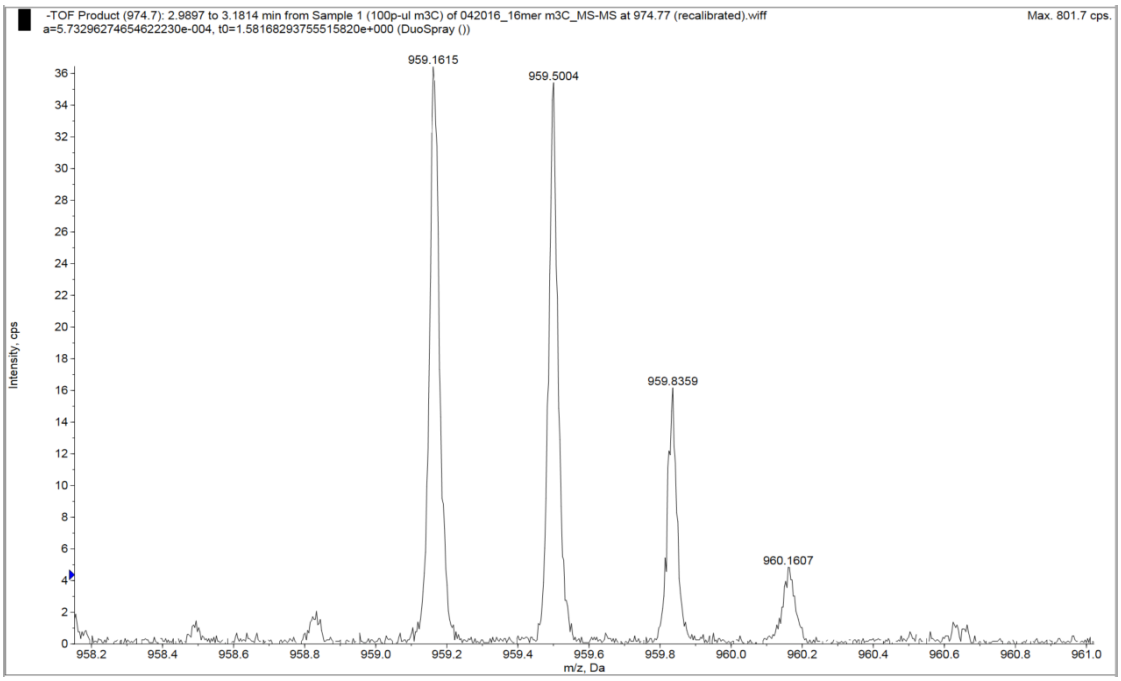

Figure S13. MS/MS fragmentation spectrum of 16mer containing m3C. The peak envelope indicates the a10-G ion at -3 charge state.

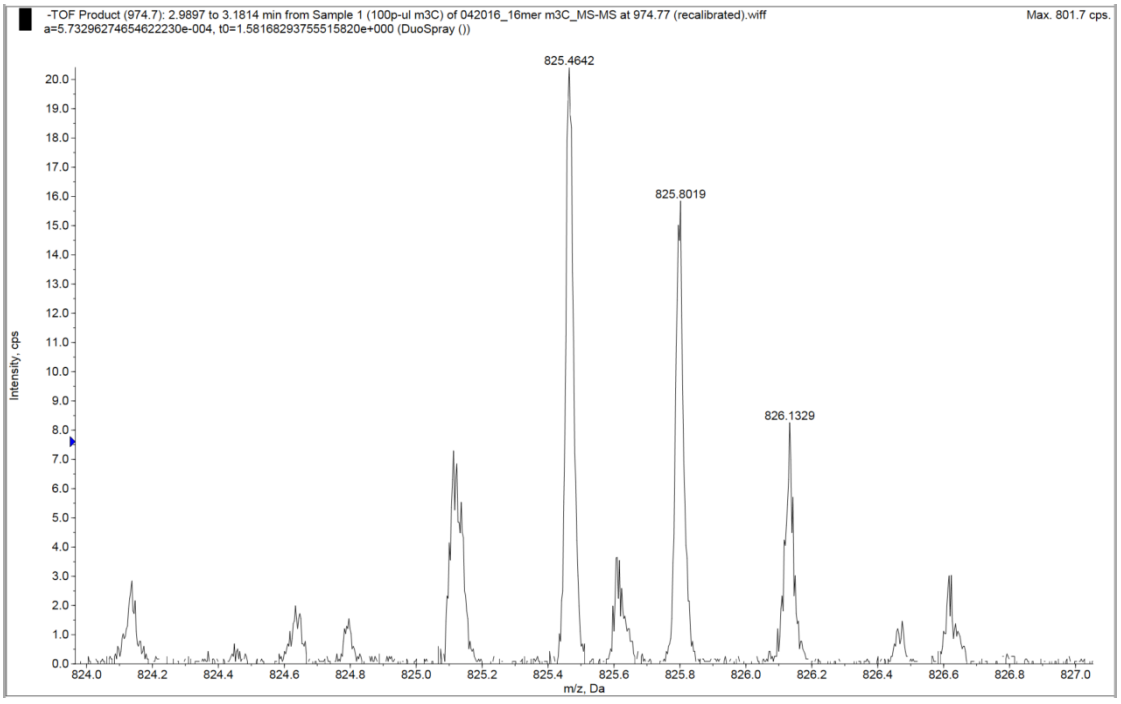

Figure S14. MS/MS fragmentation spectrum of 16mer containing m3C. The peak envelope indicates the W8 ion at -3 charge state.

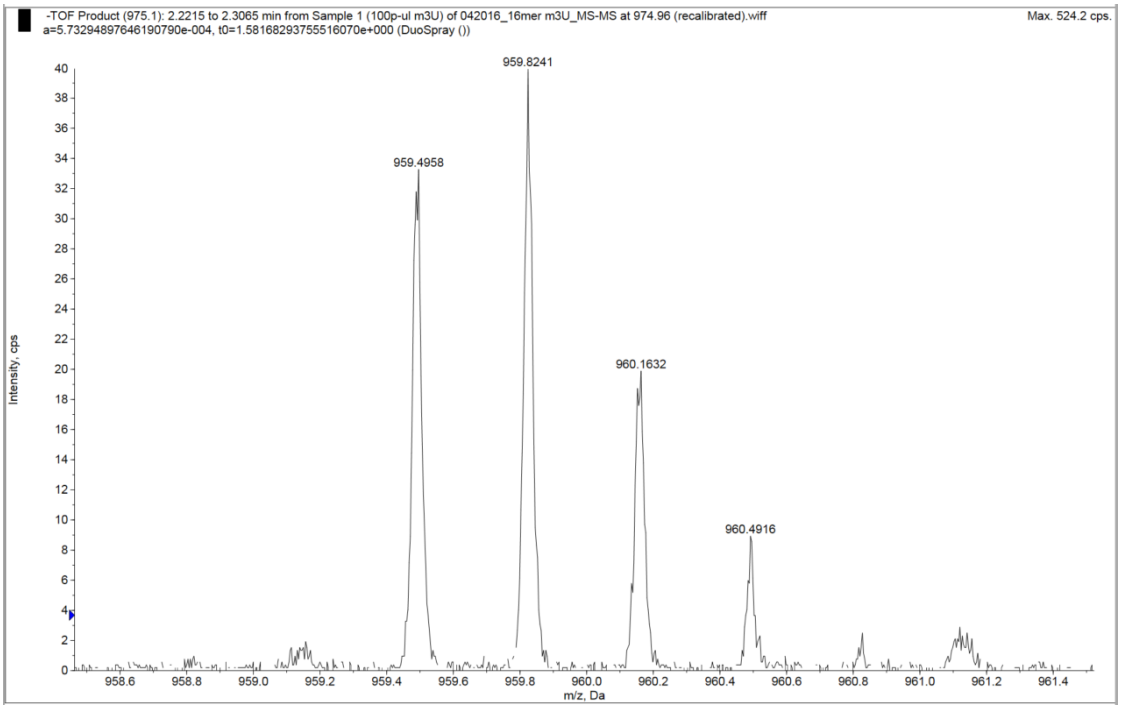

Figure S15. MS/MS fragmentation spectrum of 16mer containing m3U. The peak envelope indicates the a10-G ion at -3 charge state.

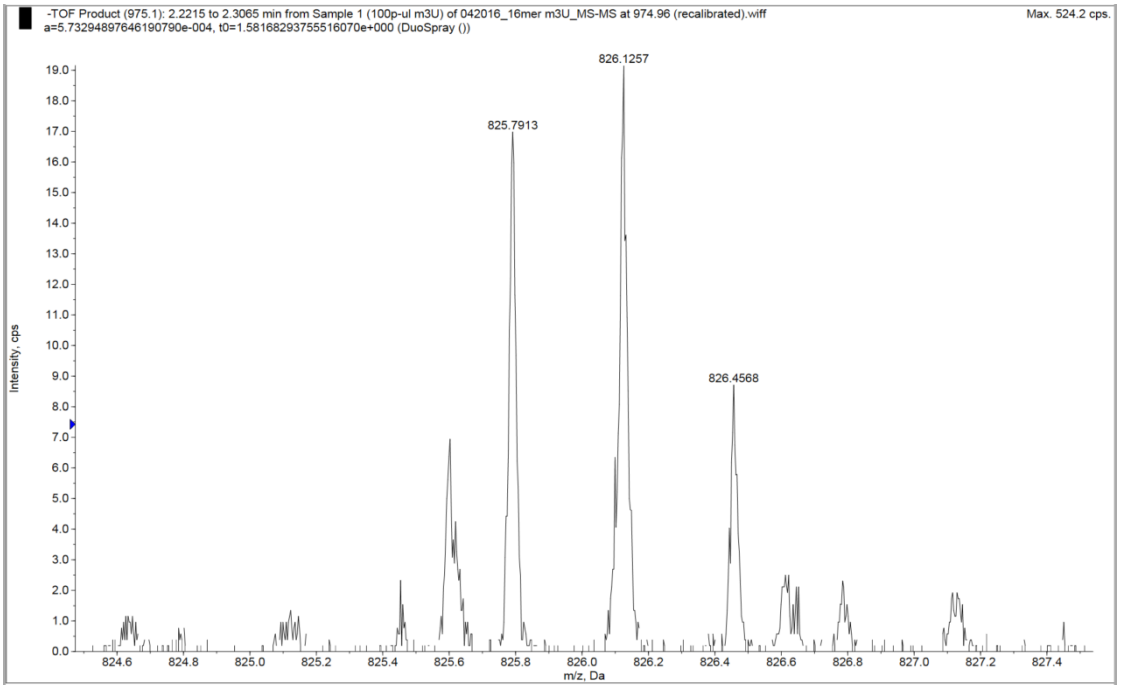

Figure S16. MS/MS fragmentation spectrum of 16mer containing m3U. The peak envelope indicates the W8 ion at -3 charge state.

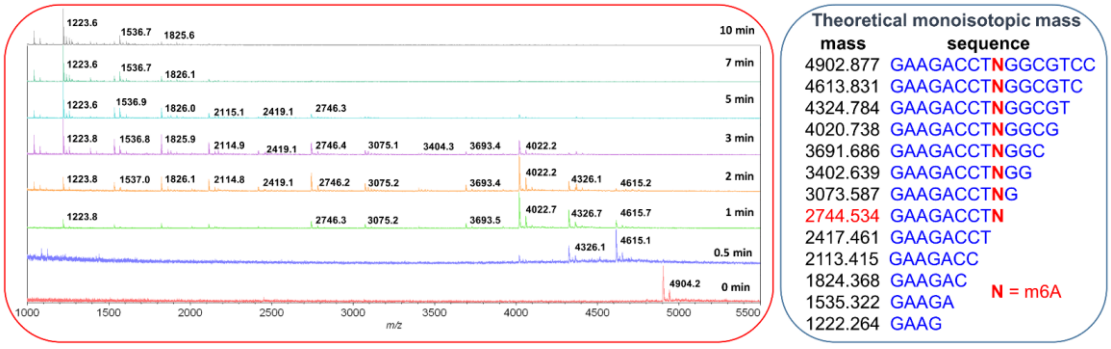

Figure S17. MALDI-TOF analyses on the exonuclease digestion of 16mer containing m6A.

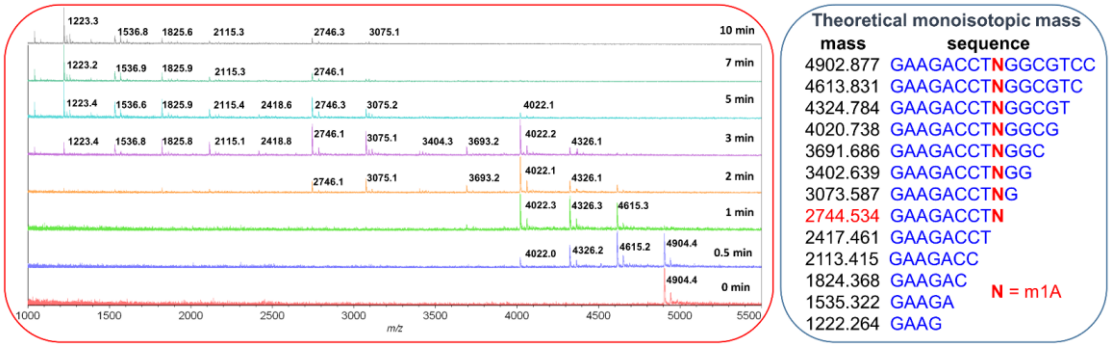

Figure S18. MALDI-TOF analyses on the exonuclease digestion of 16mer containing m1A.

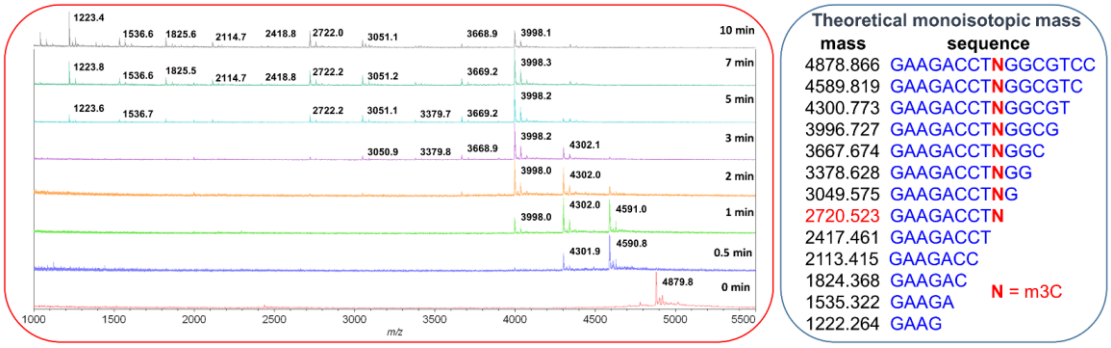

Figure S19. MALDI-TOF analyses on the exonuclease digestion of 16mer containing m3C.

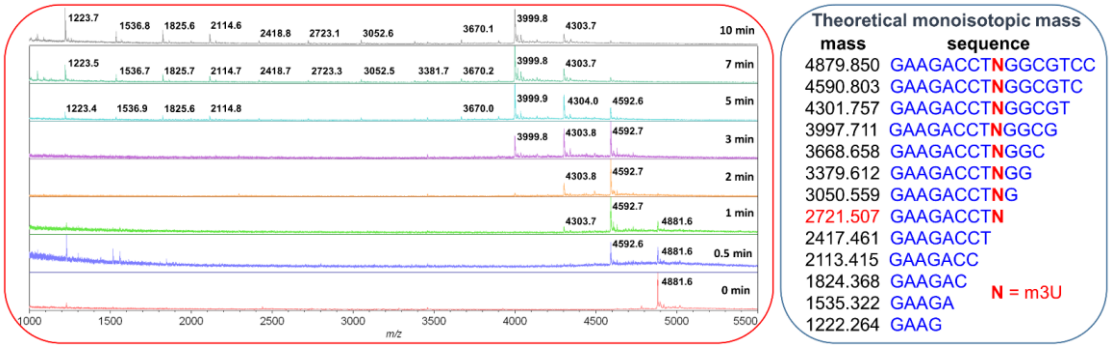

Figure S20. MALDI-TOF analyses on the exonuclease digestion of 16mer containing m3U.

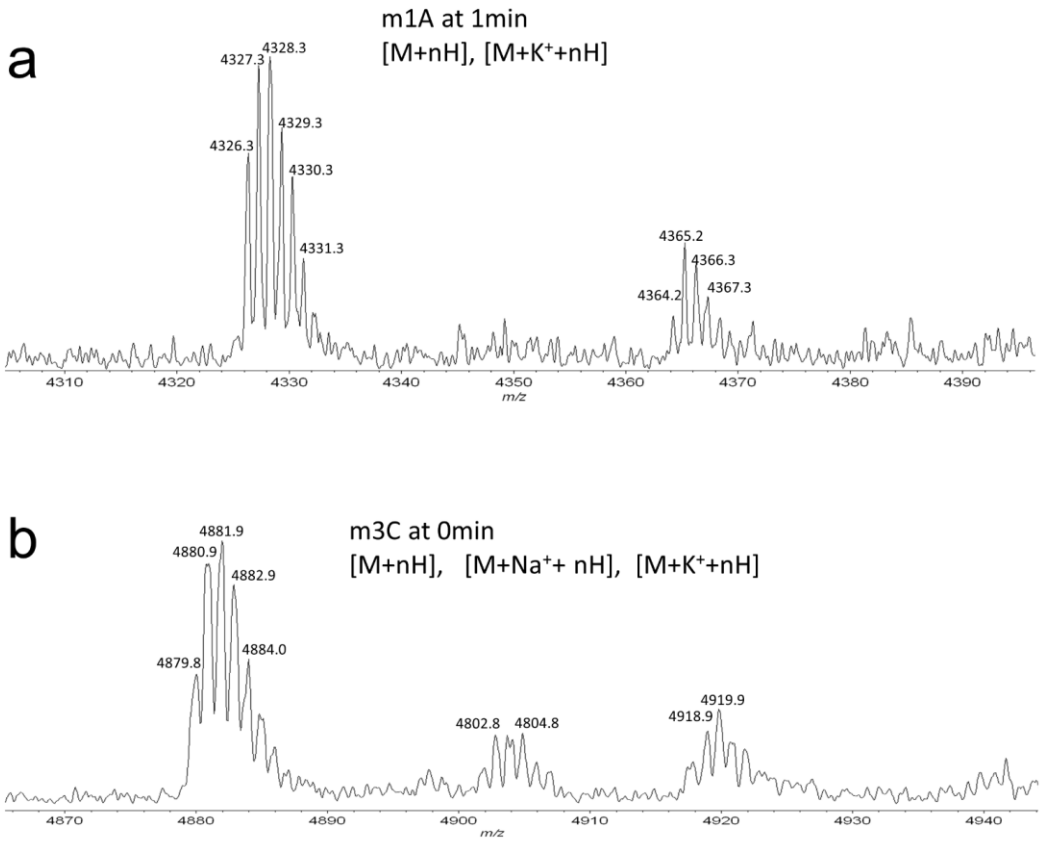

Figure S21. Detailed MALDI-TOF spectra of oligonucleotides in the exonuclease digestion assays. a) m1A and b) m3C.

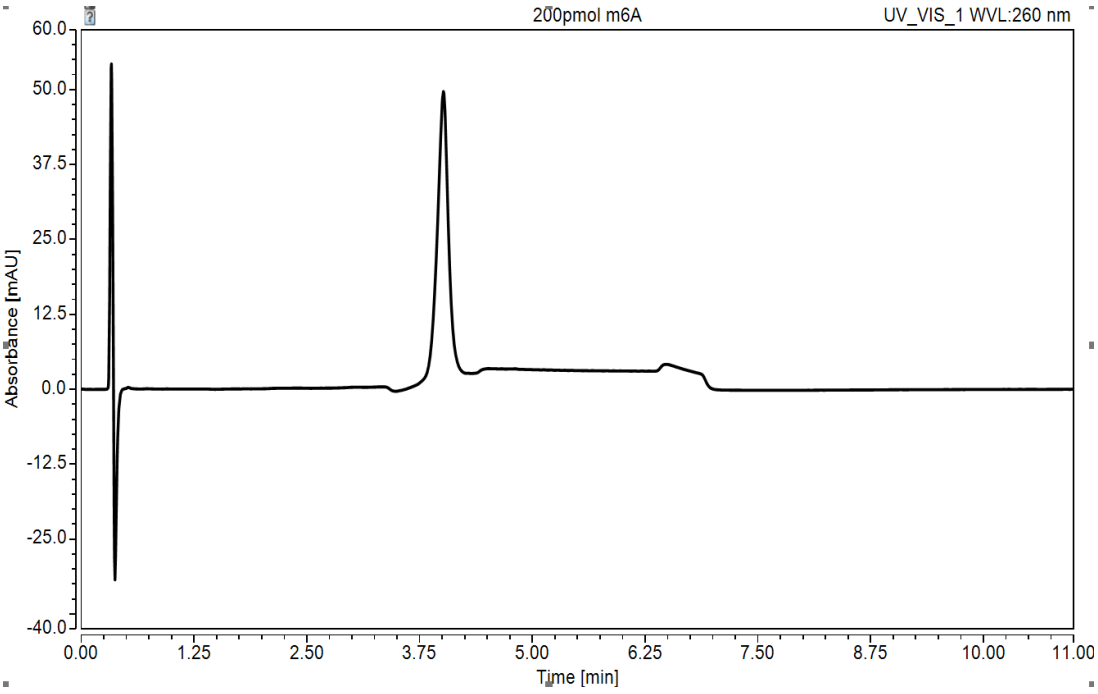

Figure S22. HPLC analysis of 16mer m6A oligonucleotide under anion exchange condition. The retention time is about 4.0 min.

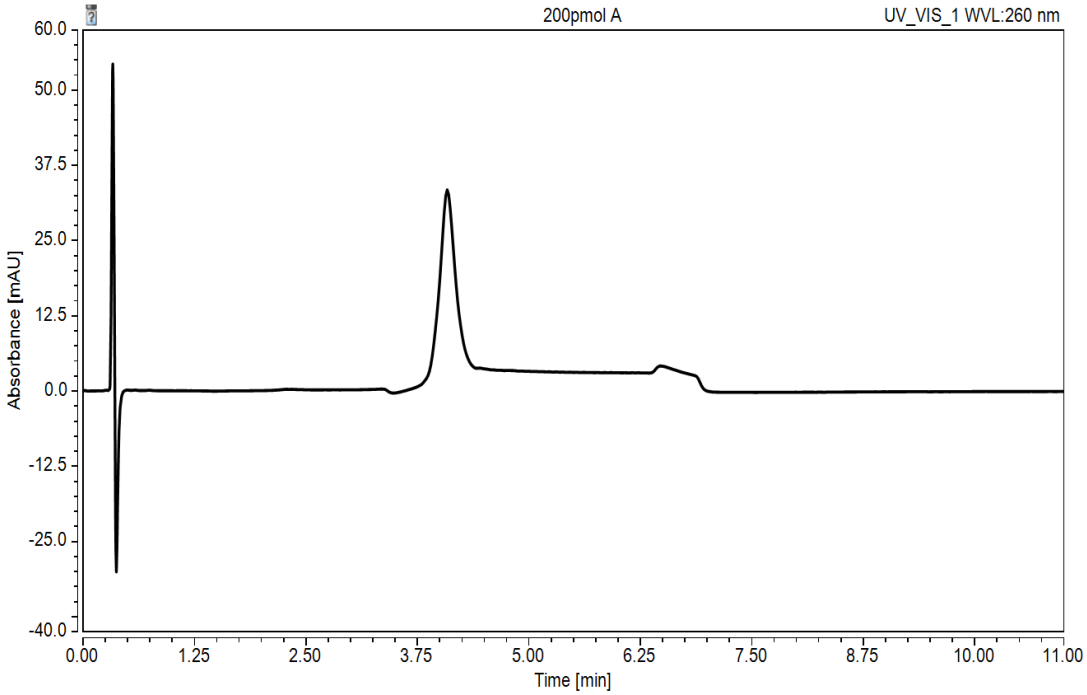

Figure S23. HPLC analysis of 16mer A oligonucleotide under anion exchange condition. The retention time is about 4.0 min.

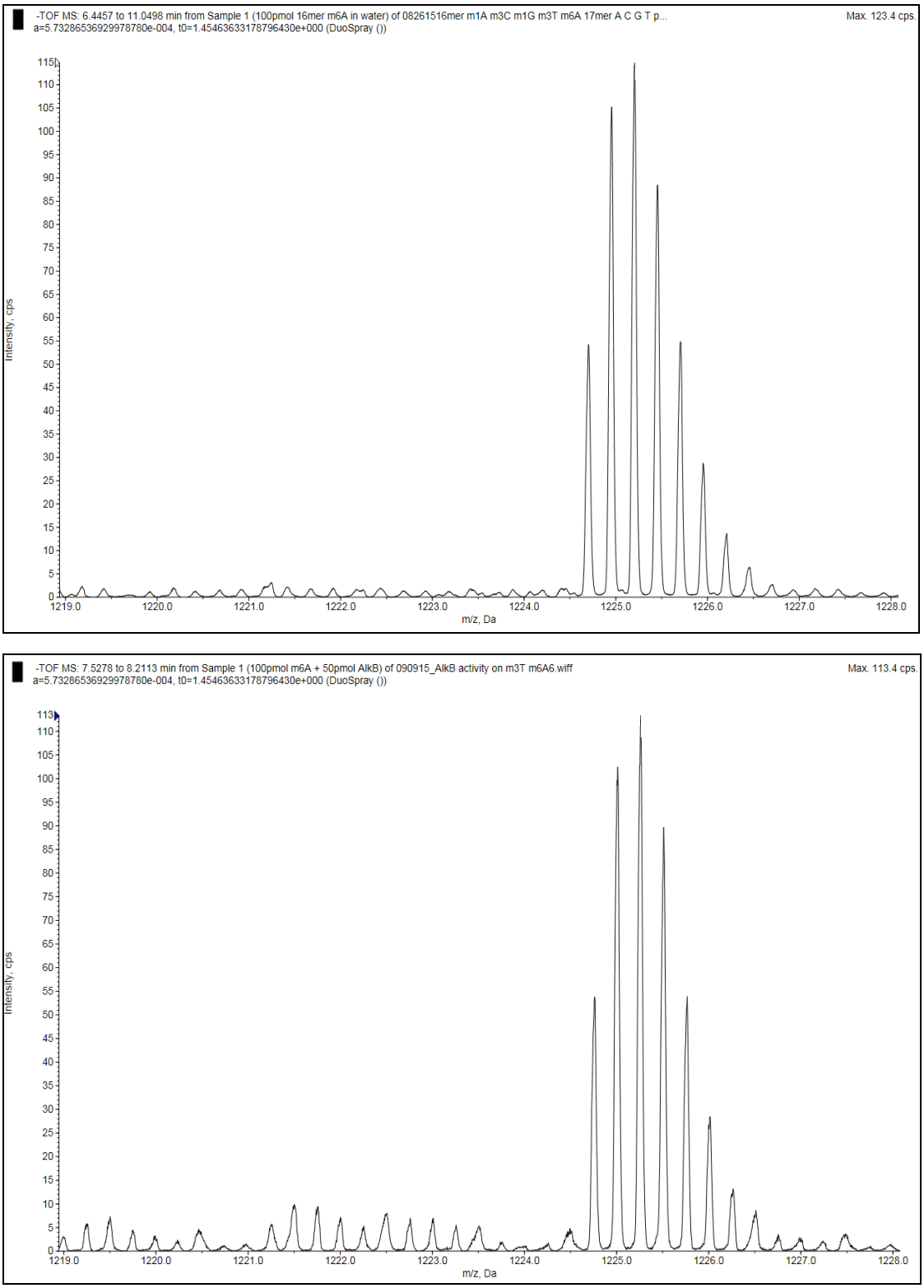

Figure S24. MS spectra of the AlkB protein repairs 16mer m6A oligonucleotide at -4 charge state. Top: 5.0 μM of m6A without addition of AlkB (peak envelope starts at

m/z 1224.7). Bottom: 5.0 μM of m6A with addition of 2.5 μM AlkB and corresponding cofactors. A very small amount of m6A (peak envelope starts at m/z 1224.7) was repaired to A (peak envelope starts at m/z 1221.2).

TABLES

Table 1. Calculated and observed monoisotopic molecular weight and m/z value of modified oligonucleotides. For m1A and m3C synthesis, the sequence of the 16mer was 5'-GAAGACCTXGGCGTCC-3', where X indicates the position of the modified bases.

Lesion or base	MW (calculated) of neutral species	*m/z* (calculated) -4 charge peak	*m/z* (observed) -4 charge peak
16mer m6A	4902.877	1224.711	1224.715
16mer m1A	4902.877	1224.711	1224.715
16mer m3C	4878.866	1218.709	1218.703
16mer m3U	4879.850	1218.955	1218.957

Table S1. Observed and predicted m/z for high resolution MS/MS fragmentation patterns displayed in Figure S5 of 16mer containing m6A. Predicted CID fragments are shown in Figure 4.

Fragment	Charge	Theoretical *m/z*	Observed *m/z*
W1	-1	306.049	306.045
W2	-2	297.044	297.043
W3	-2	449.067	449.066
W4	-2	613.593	613.593
W5	-2	758.116	758.118
W6	-3	614.759	614.753
W7	-3	724.443	724.441
W8	-3	833.468	833.467
W9	-3	934.816	934.816
W10	-3	1031.165	1031.170
W11	-4	845.384	845.379
W12	-4	923.648	923.645
W13	-4	1005.911	1005.913
W14	-5	867.139	867.131
W15	-5	929.750	929.751
a2-A	-1	426.081	426.079
a3-A	-2	369.066	369.064
a4-G	-2	525.594	525.596
a5-A	-2	690.121	690.119
a6-C	-2	846.650	846.649
a7-C	-3	660.446	660.446
a9-X	-3	858.143	858.145
a10-G	-3	967.168	967.167
a11-G	-3	1076.852	1076.851
a12-C	-4	889.650	889.650
a13-G	-4	961.912	961.911

Table S2. Observed and predicted m/z for high resolution MS/MS fragmentation patterns displayed in Figure S6 of 16mer containing m1A. Predicted CID fragments are shown in Figure 4.

Fragment	Charge	Theoretical *m/z*	Observed *m/z*
W1	-1	306.049	306.047
W2	-2	297.044	297.041
W3	-2	449.067	449.065
W4	-2	613.593	613.592
W5	-2	758.116	758.115
W6	-3	614.759	614.759
W7	-3	724.443	724.442
W8	-3	833.468	833.467
W9	-3	934.816	934.815
W10	-3	1031.165	1031.162
W11	-4	845.384	845.383
W12	-4	923.648	923.645
W13	-4	1005.911	1005.909
W14	-5	867.139	867.134
W15	-5	929.750	929.750
a2-A	-1	426.081	426.078
a3-A	-2	369.066	369.063
a4-G	-2	525.594	525.594
a5-A	-2	690.121	690.120
a6-C	-2	846.650	846.646
a7-C	-2	991.173	991.174
a8-T	-2	1135.696	1135.701
a9-X	-3	858.143	858.148
a10-G	-3	967.168	967.167
a11-G	-3	1076.852	1076.850
a12-C	-4	889.650	889.647
a13-G	-4	961.912	961.907

Table S3. Observed and predicted m/z for high resolution MS/MS fragmentation patterns displayed in Figure S7 of 16mer containing m3C. Predicted CID fragments are shown in Figure 4.

Fragment	Charge	Theoretical *m/z*	Observed *m/z*
W1	-1	306.049	306.049
W2	-1	595.095	595.095
W3	-2	449.067	449.065
W4	-2	613.593	613.593
W5	-2	758.116	758.118
W6	-2	922.643	922.644
W7	-3	724.443	724.444
W8	-3	825.464	825.464
W9	-3	926.813	926.816
W10	-4	767.119	767.117
W11	-4	839.381	839.380
W12	-4	917.645	917.647
W13	-4	999.908	999.911
W14	-4	1078.173	1078.170
W15	-4	1156.437	1156.437
a2-A	-1	426.081	426.081
a3-A	-1	739.139	739.140
a4-G	-2	525.594	525.593
a5-A	-2	690.121	690.121
a6-C	-2	846.650	846.652
a7-C	-2	991.173	991.177
a8-T	-2	1135.696	1135.701
a9-X	-3	858.143	858.144
a10-G	-3	959.164	959.162
a11-G	-3	1068.848	1068.851
a12-C	-4	883.647	883.650
a13-G	-4	955.909	955.907

Table S4. Observed and predicted m/z for high resolution MS/MS fragmentation patterns displayed in Figure S8 of 16mer containing m3U. Predicted CID fragments are shown in Figure 4.

Fragment	Charge	Theoretical *m/z*	Observed *m/z*
W1	-1	306.049	306.047
W2	-1	595.095	595.092
W3	-2	449.067	449.066
W4	-2	613.593	613.593
W5	-2	758.116	758.114
W6	-3	614.759	614.758
W7	-3	724.443	724.444
W8	-3	825.792	825.791
W9	-3	927.141	927.141
W10	-4	767.365	767.362
W11	-4	839.627	839.627
W12	-4	917.891	917.887
W13	-4	1000.154	1000.154
W14	-4	1078.419	1078.414
W15	-5	925.145	925.139
a2-A	-1	426.081	426.079
a3-A	-1	739.139	739.137
a4-G	-2	525.594	525.593
a5-A	-2	690.121	690.118
a6-C	-2	846.650	846.648
a7-C	-2	991.173	991.172
a8-T	-2	1135.696	1135.692
a9-X	-2	1287.719	1287.711
a10-G	-3	959.492	959.496
a11-G	-3	1069.176	1069.173
a12-C	-4	883.893	883.889
a13-G	-4	956.155	956.152

ACKNOWLEDGMENTS

The authors want to thank the RI-INBRE program, its director Prof. Zahir Shaikh and the staff Ms. Kim Andrews and Ms. Patricia Murray for their kind help. We also want to thank Prof. John M. Essigmann, Prof. Bingfang Yan, Prof. David Rowley, and Ms. Kerri Bradshaw for helpful discussions. This work was supported by an Institutional Development Award from the National Institute of General Medical Sciences of the National Institutes of Health under grant number 2 P20 GM103430, 1R15CA213042-01 (to D. L.), and a Medical Re-search Funds grant from the Rhode Island foundation (to D. L.). This work was in part supported by the National Institutes of Health grant CA098296 (to B.P.C.).

REFERENCES

(1) Caruthers, M. H. Gene Synthesis Machines: DNA Chemistry and Its Uses. *Science* **1985**, *230* (4723), 281–285.

(2) Beaucage, S. L.; Caruthers, M. H. Deoxynucleoside Phosphoramidites—A New Class of Key Intermediates for Deoxypolynucleotide Synthesis. *Tetrahedron Letters* **1981**, *22* (20), 1859–1862. https://doi.org/10.1016/S0040-4039(01)90461-7.

(3) Matteucci, M. D.; Caruthers, M. H. Synthesis of Deoxyoligonucleotides on a Polymer Support. *Journal of the American Chemical Society* **1981**, *103* (11), 3185–3191. https://doi.org/10.1021/ja00401a041.

(4) Leumann, C. J. DNA Analogues: From Supramolecular Principles to Biological Properties. *Bioorg. Med. Chem.* **2002**, *10* (4), 841–854.

(5) Beaucage, S. L.; Iyer, R. P. Advances in the Synthesis of Oligonucleotides by the Phosphoramidite Approach. *Tetrahedron* **1992**, *48* (12), 2223–2311. https://doi.org/10.1016/S0040-4020(01)88752-4.

(6) Petersen, M.; Wengel, J. LNA: A Versatile Tool for Therapeutics and Genomics. *Trends Biotechnol.* **2003**, *21* (2), 74–81. https://doi.org/10.1016/S0167-7799(02)00038-0.

(7) De Mesmaeker, A.; Altmann, K. H.; Waldner, A.; Wendeborn, S. Backbone Modifications in Oligonucleotides and Peptide Nucleic Acid Systems. *Curr. Opin. Struct. Biol.* **1995**, *5* (3), 343–355.

(8) *Antisense Drug Technology: Principles, Strategies, and Applications*, 2nd ed.; Crooke, S. T., Ed.; CRC Press: Boca Raton, 2008.

(9) Bonilla, J. V. *Handbook of Analysis of Oligonucleotides and Related Products.*; CRC Press: Place of publication not identified, 2016.

(10) Chen, F.; Tang, Q.; Bian, K.; Humulock, Z. T.; Yang, X.; Jost, M.; Drennan, C. L.; Essigmann, J. M.; Li, D. Adaptive Response Enzyme AlkB Preferentially Repairs 1-Methylguanine and 3-Methylthymine Adducts in Double-Stranded DNA. *Chem. Res. Toxicol.* **2016**, *29* (4), 687–693. https://doi.org/10.1021/acs.chemrestox.5b00522.

(11) Jain, V.; Hilton, B.; Lin, B.; Jain, A.; MacKerell, A. D.; Zou, Y.; Cho, B. P. Structural and Thermodynamic Insight into Escherichia Coli UvrABC-Mediated Incision of Cluster Diacetylaminofluorene Adducts on the NarI Sequence. *Chem. Res. Toxicol.* **2013**, *26* (8), 1251–1262. https://doi.org/10.1021/tx400186v.

(12) Jain, V.; Vaidyanathan, V. G.; Patnaik, S.; Gopal, S.; Cho, B. P. Conformational Insights into the Lesion and Sequence Effects for Arylamine-Induced Translesion DNA Synthesis: 19F NMR, Surface Plasmon Resonance, and Primer Kinetic Studies. *Biochemistry* **2014**, *53* (24), 4059–4071. https://doi.org/10.1021/bi5003212.

(13) Sandineni, A.; Lin, B.; MacKerell, A. D.; Cho, B. P. Structure and Thermodynamic Insights on Acetylaminofluorene-Modified Deletion DNA Duplexes as Models for Frameshift Mutagenesis. *Chem. Res. Toxicol.* **2013**, *26* (6), 937–951. https://doi.org/10.1021/tx400116n.

(14) Jain, N.; Li, Y.; Zhang, L.; Meneni, S. R.; Cho, B. P. Probing the Sequence Effects on NarI-Induced -2 Frameshift Mutagenesis by Dynamic 19F NMR, UV, and CD Spectroscopy. *Biochemistry* **2007**, *46* (46), 13310–13321. https://doi.org/10.1021/bi701386f.

(15) Singer, B.; Grunberger, D. *Molecular Biology of Mutagens and Carcinogens*; Springer US: Boston, MA, 1983.

(16) Beranek, D. T.; Weis, C. C.; Swenson, D. H. A Comprehensive Quantitative Analysis of Methylated and Ethylated DNA Using High Pressure Liquid Chromatography. *Carcinogenesis* **1980**, *1* (7), 595–606.

(17) Boiteux, S.; Laval, J. Mutagenesis by Alkylating Agents: Coding Properties for DNA Polymerase of Poly (DC) Template Containing 3-Methylcytosine. *Biochimie* **1982**, *64* (8–9), 637–641.

(18) Gomes, J. D.; Chang, C. J. Reverse-Phase High-Performance Liquid Chromatography of Chemically Modified DNA. *Anal. Biochem.* **1983**, *129* (2), 387–391.

(19) Chang, C. J.; Gomes, J. D.; Byrn, S. R. Chemical Modification of Deoxyribonucleic Acids: A Direct Study by Carbon-13 Nuclear Magnetic Resonance Spectroscopy. *The Journal of Organic Chemistry* **1983**, *48* (26), 5151–5160. https://doi.org/10.1021/jo00174a002.

(20) Kawasaki, H.; Ninomiya, S.; Yuki, H. High-Performance Liquid Chromatographic Determination of 3-Methylcytosine in Deoxyribonucleic Acid Treated with Carcinogenic Methylating Agents in Vitro and in Vivo. *Chem. Pharm. Bull.* **1985**, *33* (3), 1170–1174.

(21) Ashworth, D. J.; Baird, W. M.; Chang, C. J.; Ciupek, J. D.; Busch, K. L.; Cooks, R. G. Chemical Modification of Nucleic Acids. Methylation of Calf Thymus DNA Investigated by Mass Spectrometry and Liquid Chromatography. *Biomed. Mass Spectrom.* **1985**, *12* (7), 309–318.

(22) Kelley, M. R.; Fishel, M. L. DNA Repair Proteins as Molecular Targets for Cancer Therapeutics. *Anticancer Agents Med Chem* **2008**, *8* (4), 417–425.

(23) Messaoudi, K.; Clavreul, A.; Lagarce, F. Toward an Effective Strategy in Glioblastoma Treatment. Part I: Resistance Mechanisms and Strategies to Overcome Resistance of Glioblastoma to Temozolomide. *Drug Discov. Today* **2015**, *20* (7), 899–905. https://doi.org/10.1016/j.drudis.2015.02.011.

(24) Culp, L. A.; Dore, E.; Brown, G. M. Methylated Bases in DNA of Animal Origin. *Arch. Biochem. Biophys.* **1970**, *136* (1), 73–79.

(25) Margison, G. P.; Margison, J. M.; Montesano, R. Methylated Purines in the Deoxyribonucleic Acid of Various Syrian-Golden-Hamster Tissues after Administration of a Hepatocarcinogenic Dose of Dimethylnitrosamine. *Biochem. J.* **1976**, *157* (3), 627–634.

(26) Frei, J. V.; Swenson, D. H.; Warren, W.; Lawley, P. D. Alkylation of Deoxyribonucleic Acid in Vivo in Various Organs of C57BL Mice by the Carcinogens N-Methyl-N-Nitrosourea, N-Ethyl-N-Nitrosourea and Ethyl Methanesulphonate in Relation to Induction of Thymic Lymphoma. Some Applications of High-Pressure Liquid Chromatography. *Biochem. J.* **1978**, *174* (3), 1031–1044.

(27) Faustman, E. M.; Goodman, J. I. A Method for the Rapid Quantitation of Methylated Hepatic DNA-Purines Using High Pressure Liquid Chromatography. *J Pharmacol Methods* **1980**, *4* (4), 305–312.

(28) Beranek, D. T.; Heflich, R. H.; Kodell, R. L.; Morris, S. M.; Casciano, D. A. Correlation between Specific DNA-Methylation Products and Mutation Induction at the HGPRT Locus in Chinese Hamster Ovary Cells. *Mutat. Res.* **1983**, *110* (1), 171–180.

(29) Faustman-Watts, E. M.; Goodman, J. I. DNA-Purine Methylation in Hepatic Chromatin Following Exposure to Dimethylnitrosamine or Methylnitrosourea. *Biochem. Pharmacol.* **1984**, *33* (4), 585–590.

(30) Trewick, S. C.; Henshaw, T. F.; Hausinger, R. P.; Lindahl, T.; Sedgwick, B. Oxidative Demethylation by Escherichia Coli AlkB Directly Reverts DNA Base

Damage. *Nature* **2002**, *419* (6903), 174–178. https://doi.org/10.1038/nature00908.

(31) Falnes, P. Ø.; Johansen, R. F.; Seeberg, E. AlkB-Mediated Oxidative Demethylation Reverses DNA Damage in Escherichia Coli. *Nature* **2002**, *419* (6903), 178–182. https://doi.org/10.1038/nature01048.

(32) Fedeles, B. I.; Singh, V.; Delaney, J. C.; Li, D.; Essigmann, J. M. The AlkB Family of Fe(II)/α-Ketoglutarate-Dependent Dioxygenases: Repairing Nucleic Acid Alkylation Damage and Beyond. *J. Biol. Chem.* **2015**, *290* (34), 20734–20742. https://doi.org/10.1074/jbc.R115.656462.

(33) Sedgwick, B.; Bates, P. A.; Paik, J.; Jacobs, S. C.; Lindahl, T. Repair of Alkylated DNA: Recent Advances. *DNA Repair (Amst.)* **2007**, *6* (4), 429–442. https://doi.org/10.1016/j.dnarep.2006.10.005.

(34) Delaney, J. C.; Essigmann, J. M. Mutagenesis, Genotoxicity, and Repair of 1-Methyladenine, 3-Alkylcytosines, 1-Methylguanine, and 3-Methylthymine in AlkB Escherichia Coli. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101* (39), 14051–14056. https://doi.org/10.1073/pnas.0403489101.

(35) Pataillot-Meakin, T.; Pillay, N.; Beck, S. 3-Methylcytosine in Cancer: An Underappreciated Methyl Lesion? *Epigenomics* **2016**, *8* (4), 451–454. https://doi.org/10.2217/epi.15.121.

(36) Delaney, J. C.; Essigmann, J. M. Biological Properties of Single Chemical-DNA Adducts: A Twenty Year Perspective. *Chem. Res. Toxicol.* **2008**, *21* (1), 232–252. https://doi.org/10.1021/tx700292a.

(37) Zeng, Y.; Wang, Y. Facile Formation of an Intrastrand Cross-Link Lesion between Cytosine and Guanine upon Pyrex-Filtered UV Light Irradiation of d(^{Br}CG) and Duplex DNA Containing 5-Bromocytosine. *Journal of the American Chemical Society* **2004**, *126* (21), 6552–6553. https://doi.org/10.1021/ja049760q.

(38) Cao, H.; Jiang, Y.; Wang, Y. Stereospecific Synthesis and Characterization of Oligodeoxyribonucleotides Containing an N^2-(1-Carboxyethyl)-2‘-Deoxyguanosine. *Journal of the American Chemical Society* **2007**, *129* (40), 12123–12130. https://doi.org/10.1021/ja072130e.

(39) Delaney, J. C.; Smeester, L.; Wong, C.; Frick, L. E.; Taghizadeh, K.; Wishnok, J. S.; Drennan, C. L.; Samson, L. D.; Essigmann, J. M. AlkB Reverses Etheno DNA Lesions Caused by Lipid Oxidation in Vitro and in Vivo. *Nat. Struct. Mol. Biol.* **2005**, *12* (10), 855–860. https://doi.org/10.1038/nsmb996.

(40) Li, D.; Delaney, J. C.; Page, C. M.; Chen, A. S.; Wong, C.; Drennan, C. L.; Essigmann, J. M. Repair of DNA Alkylation Damage by the Escherichia Coli Adaptive Response Protein AlkB as Studied by ESI-TOF Mass Spectrometry. *J. Nucleic Acids* **2010**, *2010*, 369434. https://doi.org/10.4061/2010/369434.

(41) Gao, L.; Zhang, L.; Cho, B. P.; Chiarelli, M. P. Sequence Verification of Oligonucleotides Containing Multiple Arylamine Modifications by Enzymatic Digestion and Liquid Chromatography Mass Spectrometry (LC/MS). *J. Am. Soc. Mass Spectrom.* **2008**, *19* (8), 1147–1155. https://doi.org/10.1016/j.jasms.2008.04.034.

(42) Zhang, L. K.; Gross, M. L. Matrix-Assisted Laser Desorption/Ionization Mass Spectrometry Methods for Oligodeoxynucleotides: Improvements in Matrix, Detection Limits, Quantification, and Sequencing. *J. Am. Soc. Mass Spectrom.* **2000**, *11* (10), 854–865. https://doi.org/10.1016/S1044-0305(00)00161-6.

(43) Redon, S.; Bombard, S.; Elizondo-Riojas, M.-A.; Chottard, J.-C. Platinum Cross-Linking of Adenines and Guanines on the Quadruplex Structures of the AG3(T2AG3)3 and (T2AG3)4 Human Telomere Sequences in Na+ and K+ Solutions. *Nucleic Acids Res.* **2003**, *31* (6), 1605–1613.

(44) Wang, B.; Sayer, J. M.; Yagi, H.; Frank, H.; Seidel, A.; Jerina, D. M. Facile Interstrand Migration of the Hydrocarbon Moiety of a Dibenzo[a,l]Pyrene 11,12-Diol 13,14-Epoxide Adduct at N2 of Deoxyguanosine in a Duplex Oligonucleotide. *J. Am. Chem. Soc.* **2006**, *128* (31), 10079–10084. https://doi.org/10.1021/ja0608038.

(45) Tretyakova, N.; Matter, B.; Ogdie, A.; Wishnok, J. S.; Tannenbaum, S. R. Locating Nucleobase Lesions within DNA Sequences by MALDI-TOF Mass Spectral Analysis of Exonuclease Ladders. *Chem. Res. Toxicol.* **2001**, *14* (8), 1058–1070.

(46) Engel, J. D. Mechanism of the Dimroth Rearrangement in Adenosine. *Biochem. Biophys. Res. Commun.* **1975**, *64* (2), 581–586.

(47) Leutzinger, E. E.; Miller, P. S.; Kan, L.-S. Studies on the Hydrolysis of 3-Methyl-2′-Deoxycytidine in Aqueous Solution A Synthesis of 3-Methyl-2′-Deoxyuridine. *Biochimica et Biophysica Acta (BBA) - Gene Structure and*

Expression **1982**, *697* (2), 243–251. https://doi.org/10.1016/0167-4781(82)90083-5.

(48) Frick, L. E. *The Versatile E. Coli Adaptive Response Protein AlkB Mitigates Toxicity and Mutagenicity of Etheno-, Ethano-, and Methyl-Modified Bases in Vivo.*; PhD Thesis: Massachusetts Institute of Technology, 2007.

(49) Li, D.; Fedeles, B. I.; Shrivastav, N.; Delaney, J. C.; Yang, X.; Wong, C.; Drennan, C. L.; Essigmann, J. M. Removal of N-Alkyl Modifications from N(2)-Alkylguanine and N(4)-Alkylcytosine in DNA by the Adaptive Response Protein AlkB. *Chem. Res. Toxicol.* **2013**, *26* (8), 1182–1187. https://doi.org/10.1021/tx400096m.

(50) Li, D.; Delaney, J. C.; Page, C. M.; Yang, X.; Chen, A. S.; Wong, C.; Drennan, C. L.; Essigmann, J. M. Exocyclic Carbons Adjacent to the N6 of Adenine Are Targets for Oxidation by the Escherichia Coli Adaptive Response Protein AlkB. *J. Am. Chem. Soc.* **2012**, *134* (21), 8896–8901. https://doi.org/10.1021/ja3010094.

(51) Singh, V.; Fedeles, B. I.; Li, D.; Delaney, J. C.; Kozekov, I. D.; Kozekova, A.; Marnett, L. J.; Rizzo, C. J.; Essigmann, J. M. Mechanism of Repair of Acrolein- and Malondialdehyde-Derived Exocyclic Guanine Adducts by the α-Ketoglutarate/Fe(II) Dioxygenase AlkB. *Chem. Res. Toxicol.* **2014**, *27* (9), 1619–1631. https://doi.org/10.1021/tx5002817.

(52) Basiri, B.; Bartlett, M. G. LC-MS of Oligonucleotides: Applications in Biomedical Research. *Bioanalysis* **2014**, *6* (11), 1525–1542. https://doi.org/10.4155/bio.14.94.

(53) van Dongen, W. D.; Niessen, W. M. A. Bioanalytical LC-MS of Therapeutic Oligonucleotides. *Bioanalysis* **2011**, *3* (5), 541–564. https://doi.org/10.4155/bio.11.8.

(54) Lin, Z. J.; Li, W.; Dai, G. Application of LC-MS for Quantitative Analysis and Metabolite Identification of Therapeutic Oligonucleotides. *J Pharm Biomed Anal* **2007**, *44* (2), 330–341. https://doi.org/10.1016/j.jpba.2007.01.042.

(55) Schröder, A. S.; Steinbacher, J.; Steigenberger, B.; Gnerlich, F. A.; Schiesser, S.; Pfaffeneder, T.; Carell, T. Synthesis of a DNA Promoter Segment Containing All Four Epigenetic Nucleosides: 5-Methyl-, 5-Hydroxymethyl-, 5-Formyl-, and 5-Carboxy-2'-Deoxycytidine. *Angew. Chem. Int. Ed. Engl.* **2014**, *53* (1), 315–318. https://doi.org/10.1002/anie.201308469.

(56) Münzel, M.; Globisch, D.; Trindler, C.; Carell, T. Efficient Synthesis of 5-Hydroxymethylcytosine Containing DNA. *Org. Lett.* **2010**, *12* (24), 5671–5673. https://doi.org/10.1021/ol102408t.

CHAPTER 2

ADAPTIVE RESPONSE ENZYME ALKB PREFERENTIALLY REPAIRS 1-METHYLGUANINE AND 3-METHYLTHYMINE ADDUCTS IN DOUBLE-STRANDED DNA

[PUBLISHED ON CHEMICAL RESEARCH IN TOXICOLOGY 2016, 29, 687-693]

BY

Fangyi Chen,[†,◆] Qi Tang,[†,◆] Ke Bian,[†] Zachary T. Humulock,[†] Xuedong Yang,[‡] Marco Jost,[∥,○] Catherine L. Drennan,[∥,⊥,#,∇] John M. Essigmann,[§,∥,#] and Deyu Li[*,†]

[†]Department of Biomedical and Pharmaceutical Sciences, College of Pharmacy, University of Rhode Island, Kingston, Rhode Island 02881, United States.

[‡]School of Pharmaceutical Science and Technology, Tianjin University, Tianjin 300072, P. R. China.

[§]Department of Biological Engineering, [∥]Department of Chemistry, and [⊥]Department of Biology, [#]Center for Environmental Health Sciences, [∇]Howard Hughes Medical Institute, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, United States.

[◆]F.C. and Q.T. contributed equally to this work.

[*]Corresponding Author

E-mail: deyuli@uri.edu

ABSTRACT

The AlkB protein is a repair enzyme that uses an α-ketoglutarate/Fe(II)-dependent mechanism to repair alkyl DNA adducts. AlkB has been reported to repair highly susceptible substrates, such as 1-methyladenine and 3-methylcytosine, more efficiently in ss-DNA than in ds-DNA. Here, we tested the repair of weaker AlkB substrates 1-methylguanine and 3-methylthymine, and found that AlkB prefers to repair them in ds-DNA. We also discovered AlkB and its human homologs, ABH2 and ABH3, are able to repair the aforementioned adducts when the adduct is present in a mismatched base pair. These observations demonstrate the strong adaptability of AlkB on repairing various adducts in different environments.

INTRODUCTION

Nuclear DNA is constantly exposed to damage from exogenous and endogenous processes, generating a variety of DNA adducts in the genome.[1–3] If those adducts occur in a single stranded (ss) DNA region, such as at a replication fork, they may cause mutations or replication blocks during DNA synthesis. If the lesions occur in a double stranded (ds) context, they may additionally change local DNA architecture, leading to an unstable DNA duplex, or possibly disrupt the proper recognition of specific sites by sequence-specific DNA binding proteins, such as transcription factors.[1,2] To avoid these adverse effects from DNA adducts, organisms have developed an array of DNA repair pathways that are able to protect cells against lesions in both ss- and ds-DNA. One of these enzymes, the adaptive response protein AlkB of *E. coli*, has been reported to repair alkyl DNA adducts, such as 1-

methyladenine (m1A) and 3-methylcytosine (m3C), in both ss- and ds-contexts, although it prefers repairing lesions in single stranded substrates.[4–9]

The AlkB protein was discovered to be a dioxygenase that uses an α-ketoglutarate/Fe(II)-dependent mechanism to oxidize the aberrant alkyl groups, ultimately restoring the undamaged DNA bases (Figure 1a).[4,5] Different homologs of the *E. coli* AlkB protein exist in prokaryotic and eukaryotic species; nine such homologs exist in human cells (ABH1-8 and FTO).[4,5] Among the nine AlkB homologs, ABH1 functions as an apyrimidinic/apurinic lyase and nucleic acid demethylase.[6,7] ABH4[8,9] and ABH7[10,11] modify protein substrates and ABH8[12–16] is a tRNA methyltransferase and hydroxylase. ABH5[17–21] and FTO[22–25] have been demonstrated to work on N6-methyladenine (m6A) in RNA or DNA. ABH2[2,26–30] and ABH3[31–34] are DNA repair enzymes; and the function of ABH6 remains to be established. Since the discovery of this class of enzymes, a variety of DNA adducts have been identified as substrates for AlkB and its mammalian homologs, ABH2 and ABH3, both *in vitro* and *in vivo*.[1,10–13] The adducts include all of the seven N-methyl lesions occurring at the Watson-Crick (W-C) base-pairing face of the four nucleobases.[13] The seven adducts include m1A, m3C, m6A, N4-methylcytosine (m4C), 1-methylguanine (m1G), N2-methylguanine (m2G), and 3-methylthymine (m3T). Among these lesions, m6A, m4C and m2G are exocyclic adducts whose structures afford the opportunity to avoid disruption of W-C base-pairing by allowing the methyl group to swivel away from the H-bond interface. By contrast, m1A/m3C/m1G/m3T have the methyl group on the nucleobase ring, which unavoidably will interfere with hydrogen-bond pairing if left unrepaired. AlkB has

also been reported to repair other DNA adducts, such as 3-ethylcytosine (e3C), N2-ethylguanine, 1,N6-ethenoadenine (εA), 3,N4-ethenocytosine (εC), 1,N2-ethenoguanine (1,N^2-εG), 1,N6-ethanoadenine (EA), 3,N4-α-hydroxyethanocytosine (HEC), 3,N4-α-hydroxypropanocytosine (HPC), N2-furan-2-yl-methylguanine, N2-tetrahydrofuran-2-yl-methylguanine, α-hydroxypropanoguanine, γ-hydroxypropanoguanine, and malondialdehydeguanine.[35–43] The substrate scope and repair efficiency of the AlkB family enzymes have been reviewed by several papers.[1,5,44,45]

EXPERIMENTAL PROCEDURES

Oligonucleotide synthesis. Sixteen-mer oligonucleotides (oligos) with the sequence 5'-GAAGACCTXGGCGTCC-3' containing the lesions at X position were made by using solid-phase phosphoramidite chemistry[36,37,40–42] on a MerMade-4 Oligonucleotide Synthesizer. The complementary 17mer oligos were synthesized with the sequence of 5'-TGGACGCCYAGGTCTTC-3', where Y represents the position incorporating the regular bases A, C, G and T. For the ss-23mer unrelated DNA, the sequence was 5'-AAAGCTTCTGCAATCAGGTTCAG-3'. The oligos were purified by reverse-phase HPLC with two solvents. Solvent A was 100 mM 1:1 triethylamine-acetic acid (TEAA) in water and solvent B was 100% acetonitrile. The column was Phenomenex DNAPac PA-100 Semi-Preparative (9 x 250 mm, 5μm). The concentration of DNA was determined by UV absorbance at 260nm. The extinction coefficient (ε) of a certain adduct is calculated as its unmodified counterpart due to the negligible variation between the values in the context of a 16-mer DNA. The oligos were characterized by HPLC-electrospray ionization (ESI) triple quadrupole-TOF

mass spectrometry (MS) (AB Sciex). Solvent A was 10mM ammonium acetate in water and solvent B was 100% acetonitrile. The column was Phenomenex Luna C 18 column (4.6 × 100 mm; 5μm). The calculated and observed monoisotopic MW and m/z value of the oligos are shown in Table S1.

Expression and purification of the AlkB, ABH2 and ABH3 proteins. The AlkB gene was cloned into a pET28a+ vector (EMD Millipore) and then transformed into *E. coli* Rosetta2(DE3)pLysS (EMD Millipore) cells for expression.[40,41] The expressed protein was purified by affinity column chromatography, HisTrap HP (GE Healthcare Life Sciences). Thrombin (Sigma-Aldrich, 0.005U/10μg protein) was used to digest His-tag containing AlkB protein overnight. Cation-exchange column chromatography, HiTrap SP HP (GE Healthcare Life Sciences) was used for further purification. The final purified protein was concentrated by Amicon® Ultra Centrifugal Filters (EMD Millipore) and stored in the AlkB storage buffer (10 mM Tris, 100 mM NaCl, 1 mM 2-mercaptoethanol, 10% glycerol, pH 8.0). The ABH2 and ABH3 proteins were purified using the same procedure, with the exception that the *E. coli* cell used for expression was BL21(DE3)pLysS (EMD Millipore) instead of Rosetta2(DE3)pLysS. The ABH2 and ABH3 proteins were stored in the ABH storage buffer (50 mM N-[Tris(hydroxymethyl)methyl]-3-aminopropanesulfonic acid, 300 mM NaCl, 10% glycerol, and 1 mM 2-mercaptoethanol, pH 8.0). Protein standard was purchased from Bio-Rad.

Enzymatic reaction. All reactions were performed at 37 ℃ for 1h in a reaction buffer containing 70 μM Fe(NH4)2(SO4)2·6H2O, 0.93 mM α-ketoglutarate, 1.86 mM ascorbic acid, and 46.5 mM HEPES (pH 8.0).[37,39,40] The reactions were quenched by

adding 10 mM EDTA followed by heating at 95 ℃ for 5 min. Samples were then analyzed by HPLC or HPLC-electrospray ionization triple quadrupole-TOF MS (AB Sciex). Typically, the purified AlkB, ABH2 and ABH3 proteins were incubated with 5 μM DNA oligos in the presence of all cofactors in a 20 μL reaction volume. The amount for these enzymes were variable on different substrates as shown in Table S3-5. For the double-stranded DNA substrates, 1.5 equivalents (7.5 μM) of the 17mer complementary oligo containing A/C/G/T opposite to the adduct was annealed by heating the mixture at 80 ℃ for 10 min and then cooled down slowly to room temperature; the rest of the reaction was under similar conditions to ss-DNA repair reaction. For the initial velocity measurements, the reactions were stopped at 0, 0.5, 1, 4, and 10min for m1A and m3C and at 0, 4, 8, and 12 min for m1G and m3T. For the reactions with excessive amount of unrelated DNA, each reaction was carried out with 5μM substrate and additional 15μM ss-23mer DNA. The corresponding concentrations of the enzymes were listed in Table S5. Each reaction was carried out in triplicate.

LC-MS analyses. Oligonucleotide analyses were performed on AB Sciex triple quadrupole-TOF mass spectrometer. ESI was conducted using a needle voltage on 4.0 kV. Nitrogen gas was used with a setting of drying 40 L/min and a heated capillary at 600 ℃. Liquid chromatographic separation was performed using a Phenomenex Luna C 18 column (4.6 × 100 mm; 5μm) at a flow rate 0.4 mL/min. Solvent A was 10mM ammonium acetate in water and solvent B was 100% acetonitrile.[40,41] A linear gradient was carried out under the following conditions: 2.0% of B for 0.5 min, 2.0 to 17.4% of B over 11 min, 17.4 to 60.0% of B over 0.1 min, 60.0% of B for 2 min, 60.0

to 2.0% of B over 0.1 min, then 2.0 % B over 3.3 min. The LC analyses were carried out with the temperature of the column oven set at 40 ℃. Data analyses were performed with the AB Sciex Analyst TF software 1.7. The relative amount of the starting material and product was quantified by integrating the peak areas of the corresponding oligos in the mass spectrum. In each reaction, we treated 16mer oligo containing the starting material had the same ionization efficiency as the 16mer product due to the negligible variation between the structures in the context of a 16-mer DNA.

Statistical analyses. All statistical analyses were carried out by IBM SPSS Statistics 23 and Microsoft Excel 2013. Statistical significance for data between two groups were performed with Student's two-tailed *t* test. P-value <0.05 was considered as statistically significant.

RESULTS

AlkB has been reported to repair strong substrates, such as m1A and m3C, more efficiently in ss-DNA than in ds-DNA.[4–9] However, the strand preference of AlkB for some of its weak substrates, such as m1G and m3T, remained unanswered. In this work, we selected four lesions to investigate the repair efficiency of AlkB in both ss- and ds-DNA (Figure 1b). Because these adducts are mutagenic to varying degrees, causing misincorporation of non-authentic pairing bases during replication, we also tested the ds-DNA conditions with all four DNA bases opposite to each adduct in the complementary strand to demonstrate AlkB's repair capacity under mismatch conditions. ABH2 and ABH3 have been reported to prefer repairing DNA adducts in ds- and ss-DNA, respectively, and were also tested on those lesions.

The ability of AlkB family enzymes to repair the four methylated bases was tested *in vitro* using a previously reported experimental procedure.[12,13] AlkB and its human homologs were purified from *E. coli* cells (Figure S3-5). Oligos containing the four adducts were synthesized by incorporating each adduct into a 16mer oligo in a site-specific manner and their purity and identity were characterized by LC-MS (Figure S6 and S8-11).[11] For each lesion, experiments were conducted both in the presence and absence of the repair enzyme with all necessary cofactors, and the reaction products were analyzed by high resolution triple quadrupole-TOF MS (Figure 2). To test repair efficiency in ds-DNA, each adduct was first annealed to 17mer complementary oligos (Figure S7 and S12-15) with all four DNA bases opposite the lesion. To avoid the interference of the complementary oligo with the 16mer starting material in the MS spectra, a 17mer complementary oligo was chosen because it has ~300 Daltons difference in MW than the 16mer oligo. The 16mer lesion-bearing oligos demonstrated a good signal in the -4 charge envelope of the MS spectra (Figure 2). To illustrate the MS analysis, the MW of the 16mer containing an m1G lesion is calculated as 4918.87 Daltons (Table S1); its monoisotopic peak (all 12C, 14N, 16O, etc.) in the -4 charge state has a calculated mass/charge (m/z) of 1228.71. Experimentally, we observed a peak at an m/z of 1228.70 (Figure 2c). The area underneath the peak envelopes of the starting material and product were used to quantify the conversion of the reaction (Figure 2).

For each lesion-containing oligo, 5μM starting material was used in the repair reaction. To accurately compare the efficiency of repair for each lesion, the concentration of each enzyme was optimized to ensure no reaction proceeded to 100%

conversion (Table S3). Acting upon an m1A lesion in ss-DNA, AlkB converted 53% of the starting material to a 16mer product containing A at the lesion site (Figure 2a and 3a). When m1A was annealed to the 17mer oligo containing complementary base T, the theoretical perfect match in a repaired state, only 22% of m1A was repaired. For m1A paired with mismatched bases, such as A/G/C, the reactions also showed lower repair ratios than in ss-DNA (Figure 3a and Table S2). Similarly, AlkB was demonstrated to repair m3C better in ss-DNA than ds-DNA with the exception that m3C:A mismatched pair provided a slightly higher repair efficiency (Figure 3b and Table S2). These results were consistent with the previous observations of AlkB's repair of m1A and m3C in ss- and ds-DNA.[4–9]

After testing the strong substrates of AlkB (m1A and m3C), the two weak substrates m1G and m3T were investigated. For m1G, 48% of the adduct was repaired in ss-DNA, but 69% was repaired when m1G was put opposite to C (Figure 2c, 2d, 3c, and Table S2). For the mismatch reactions, complementary base A (50%) and T (56%) also showed higher repair ratios than the single-stranded reaction. For m3T repair, the ss-DNA condition (12%) provided the least efficient repair among all conditions (Table S2). The perfect match of m3T with A was repaired 21% and the mismatch with C yielded 28% conversion. Because m1A and m3C are the strong substrates of the three enzymes and m1G and m3T are the weak substrates of the proteins, the concentrations of the enzymes for different reactions varied significantly from 0.08-5.0 μM in the steady state studies (Table S3). The concentrations were optimized to make sure that the repair ratios were significantly different between ss- and ds-repair. The results from m1G and m3T reactions demonstrated AlkB repairs

them more efficiently in ds-perfect match conditions than in ss-DNA in the steady state studies. To further support this conclusion, we also measured the initial velocity of the repair reactions on the adducts in ss-DNA and in the perfect match ds-DNA (Table 1). The results confirmed the preferential repair of m1G and m3T in ds-DNA, and the differences of repair efficiency between ss- and ds-conditions were statistically significant. On the other hand, the initial rates of repair on m1A and m3C were higher in ss-DNA than ds-DNA (Table 1).

The current studies on AlkB's repair of the four DNA adducts were carried out in a 16mer sequence context, which is different from the situation in cell where adducts are more likely to be present in parts per thousand or less. To study the repair of AlkB with an excessive amount of non-damaged DNA, the repair reactions were performed by adding of an extra 15 μM unrelated ss-23mer DNA oligos. This condition allowed us to decrease the lesion:normal base ratio from 1:15 to 1:84. The repair efficiency also supported the previous conclusions that AlkB prefers to repair m1G and m3T in ds-DNA and m1A and m3C in ss-DNA with the repair ratio differences statistically significant (Figure S16 and Table S5).

We also tested the repair of these four adducts by AlkB's human homologs ABH2 and ABH3. For the repair ratios of the four lesions with ABH2, the double stranded condition for the perfect match base pairs and most of the mismatch base pairs were better repaired than ss-DNA, except the m3C:T mismatched pair (54%), which demonstrated a slightly lower repair efficiency than m3C in ss-DNA (60%, Figure S1 and Table S2). For the repair of m1A and m3C by ABH3, ss-DNA repair was consistently more effective than in ds-DNA (Figure S2 and Table S2). However, we

were not able to detect the reaction products of m1G and m3T because they showed very little repair susceptibility, even when ABH3 was present at a very high concentration (5 μM, Table S3). These results were consistent with previous observations that ABH2 prefers ds repair and ABH3 prefers ss repair.[28,29,46–50]

DISCUSSION

In the above section, we demonstrated that AlkB is able to repair DNA adducts in both ss- and ds-DNA. AlkB repairs m1A and m3C better in ss-DNA than in ds-DNA for most strand conditions. However, the preference of repairing m1G and m3T is the opposite: AlkB prefers to repair these lesions in perfect match and most of the mismatch ds-DNA. AlkB has been reported to repair those four DNA adducts both *in vitro* and *in vivo*.[36,39] It has been demonstrated that m1A and m3C are strong substrates for AlkB because both are positively charged under physiological pH conditions.[35,36] In AlkB deficient cells, m1A and m3C behave as strong replication blocks: only ~12% of them could be bypassed by the replication polymerase.[36] m1A is not mutagenic and m3C is ~ 30% mutagenic with the predominant mutations being C to T and C to A. The existence of AlkB could completely repair m1A and m3C in cell, thus fully alleviating their toxicity and mutagenicity. AlkB can also efficiently repair e3C, εA and EA both *in vitro* and *in vivo*.[36–38,40] On the other hand, m1G, m3T, 1,N^2-εG and εC are weak substrates for AlkB with substantial cytotoxic and mutagenic signatures in both AlkB positive and negative cells.[36,37,43] For example, m1G and m3T are only bypassed less than 10% of the replications in the absence of the AlkB protein. The presence of AlkB only slightly increases the bypass of these two adducts. For mutagenicity, m1G is ~80% mutagenic with G to T, G to A and G to

C mutations without the repair of AlkB. Similarly, m3T is ~60% mutagenic with T to A and T to C mutations. The mutagenicity of m1G and m3T decreases with the presence of AlkB. The cellular results of AlkB's repair on those four adducts support the current observation that m1A and m3C are strong substrates of AlkB, and m1G and m3T are weak substrates. If these two weak substrates are left unrepaired in cell, they will not only strongly block replication but also cause a significant increment in the percentage and type of mutations. Based on these observations, we hypothesize that AlkB could efficiently repair strong substrates, such as m1A and m3C, in ss-DNA before they encounter a polymerase. In contrast, AlkB cannot efficiently repair weak substrates, such as m1G and m3T, when they exist in ss-DNA. AlkB may have evolved to repair the weak substrates better in a ds-DNA context. This preference would avoid mutations to a certain degree, or in combination with mismatch repair, could potentially prevent mutations pre- and post-replication.

Previously, Zhu et al. identified an active site region and key amino acid residues that are responsible for the repair of m1A by AlkB and the repair of m6A by FTO and ABH5.[51] By swapping the active site sequences between the two types of demethylases, they found the enhanced activity of not only AlkB on repairing m6A but also FTO and ABH5 on m1A. Similarly, Chen et al. achieved preferential ss-repair by ABH2 and ds-repair by ABH3 (the opposite strand preference for both of the enzymes) by swapping the recognition residues in their structures.[27] AlkB has been reported to repair different DNA adducts in both ss- and ds-DNA. Further structural analysis should provide valuable information on AlkB's strand preference of m1A/m3C and m1G/m3T. Besides the strand preference, we also demonstrated that

all three enzymes can repair lesions in ds-DNA under mismatch conditions; some of the mismatches are even better repaired than the perfect match. These observations demonstrate the strong adaptability of AlkB on repairing various adducts under different strand and sequence contexts. The current studies utilized pH 8.0 and an iron concentration that would provide a steady state measurement. However, Maciejewska et. al have shown that AlkB's repair efficiency depends on the pKa of substrates as well as optimal Fe(II) ion concentration.[35] They demonstrated the optimal repair of m3C and HPC was achieved at pH 7.5, and the optimal repair of HEC was achieved at pH 5.8. Correspondingly, the adducts would be positively charged under the above pH conditions. The preference to positively charged adducts is possibly due to the interaction with the negatively charged aspartate residue (Asp135) in the catalytic center of AlkB. On the other hand, the poor repair efficiency on m1G and m3T could be explained, at least partially, that the methylated nitrogen atoms in those two adducts do not exist in cationic form under any pH condition (not limited to physiologically relevant pH range).[35,42] The molecular and structural mechanisms of AlkB's preference on ss- or ds-DNA warrant further investigation.

FIGURES

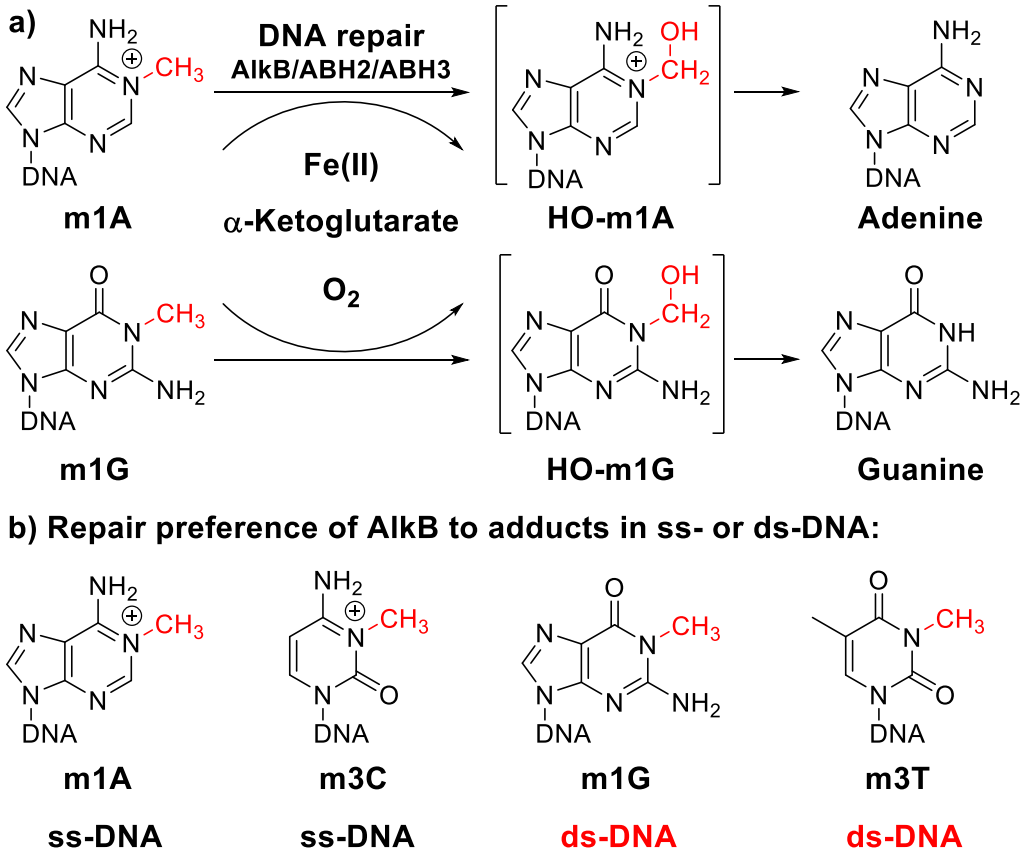

Figure 1. a) Repair mechanism of the AlkB family enzymes on alkyl DNA lesions. Adducts m1A and m1G are used here as examples to represent strong and weak substrates of AlkB, respectively. b) The four DNA adducts studied in this work and their strand preference (ss- vs ds-DNA) for AlkB to act on them.

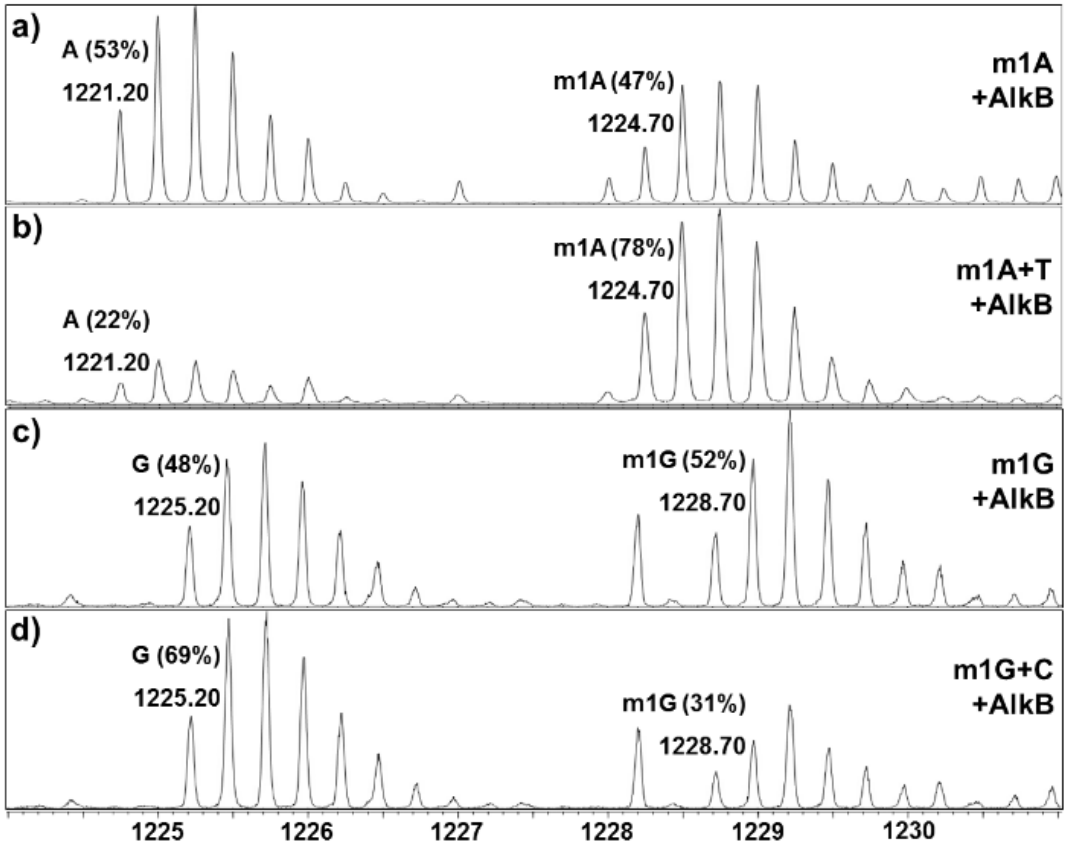

Figure 2. High resolution triple quadrupole -TOF MS analyses of AlkB repairing different alkyl adducts in ss- and ds-DNA. Data represent the starting material and product in their -4 charge envelopes, with the observed m/z values of their monoisotopic (all 12C, 14N, 16O, etc.) peaks labeled above each envelope. The percentage of the starting material and product in each reaction is also labeled above the corresponding peak envelopes. The peak near m/z 1228.20 in panel c and d is from a non-DNA impurity. a) AlkB + ss-m1A; b) AlkB + ds-m1A:T; c) AlkB + ss-m1G; and d) AlkB + ds-m1G:C.

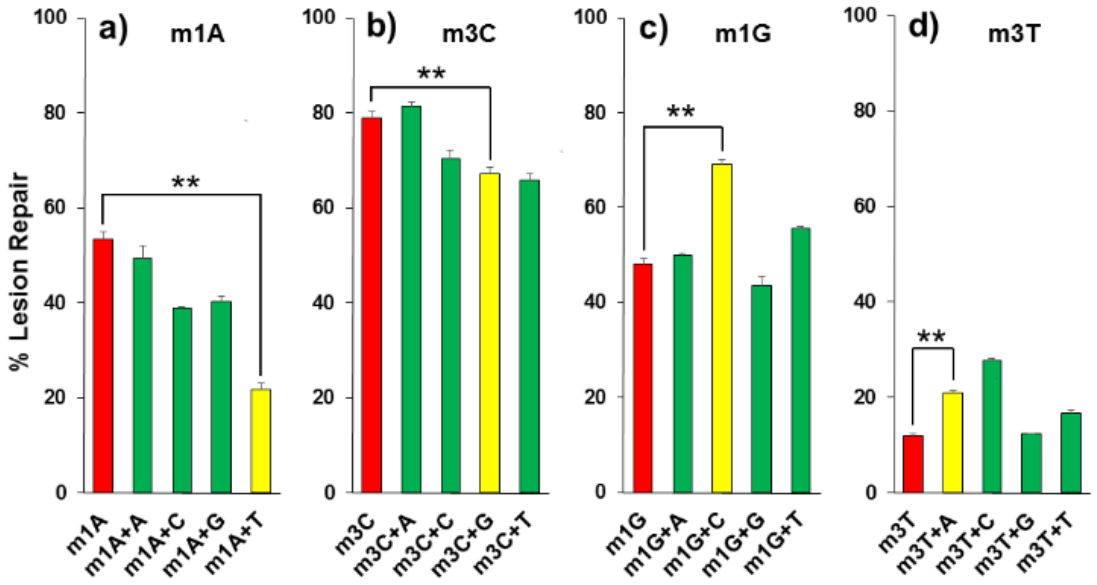

Figure 3. Repair efficiency of AlkB on different 16mer adducts in ss- and ds-DNA. For each lesion, the extent of repair in ss-DNA is colored in red, the extent of repair in ds-DNA with non-mutagenic pairing is colored in yellow, and the extent of repair in ds-DNA with mismatch bases is colored in green. Detailed information of enzyme concentrations is listed in Table S3. The error bars represent the standard deviation from triplicate experiments. The significance of the difference between ss-DNA and ds-DNA with non-mutagenic pairing was tested using the Student's two-tailed *t* test. ** indicates p-value <0.01. a) m1A; b) m3C; c) m1G; and d) m3T.

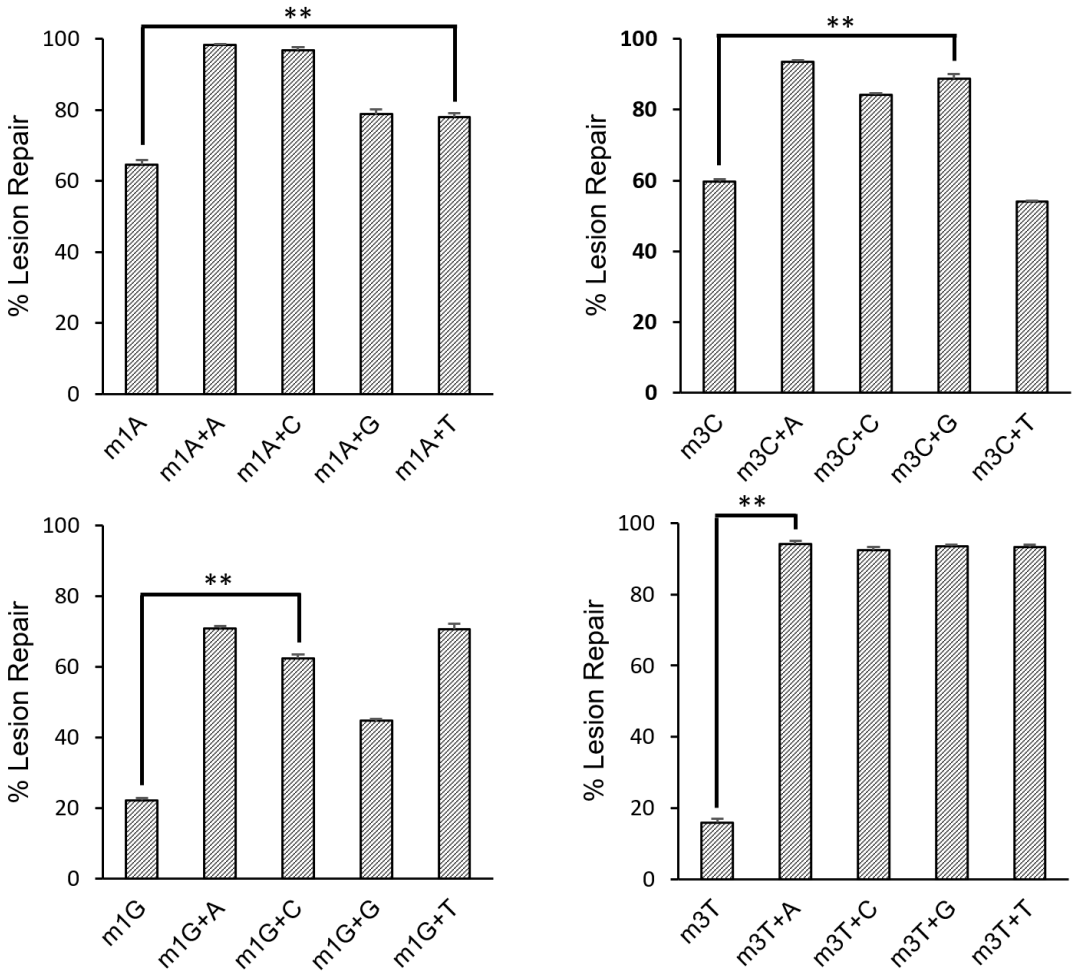

Figure S1. Repair efficiency of ABH2 on different adducts in ss- and ds-DNA. The error bars represent the standard deviation from triplicate experiments. The significance of the difference between ss-DNA and ds-DNA with non-mutagenic pairing was tested using the Student's two-tailed t test. ** indicates p-value <0.01.

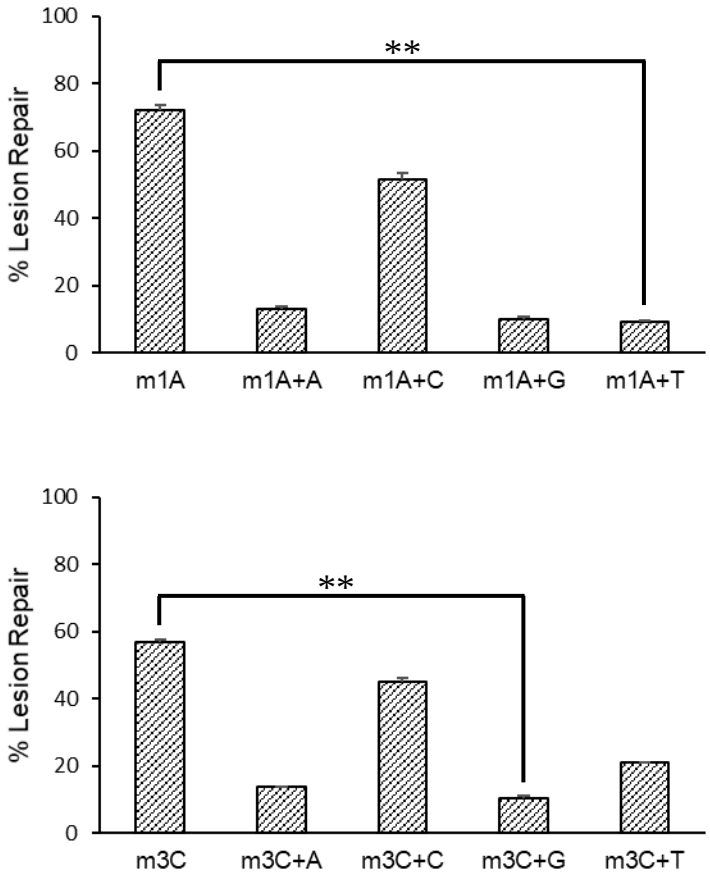

Figure S2. Repair efficiency of ABH3 on different adducts in ss- and ds-DNA. The error bars represent the standard deviation from triplicate experiments. The significance of the difference between ss-DNA and ds-DNA with non-mutagenic pairing was tested using the Student's two-tailed t test. ** indicates p-value <0.01.

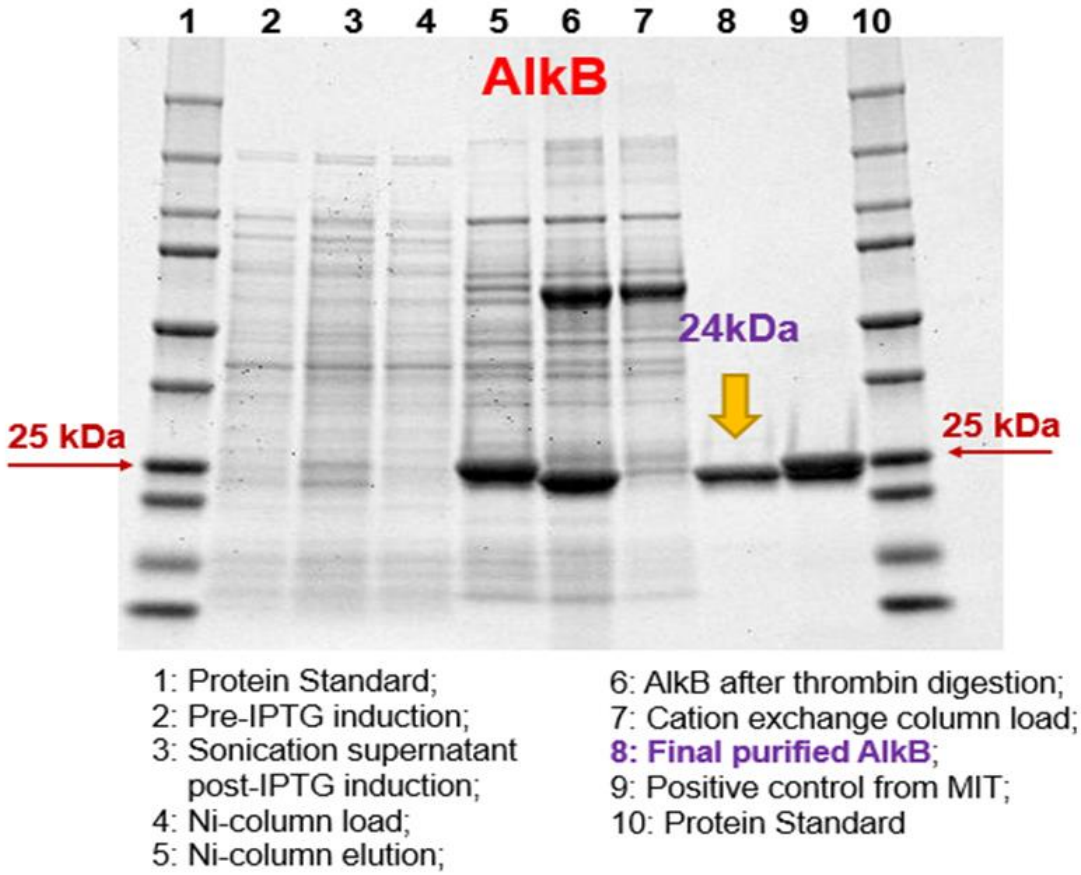

Figure S3. Polyacrylamide gel electrophoresis of purified AlkB protein.

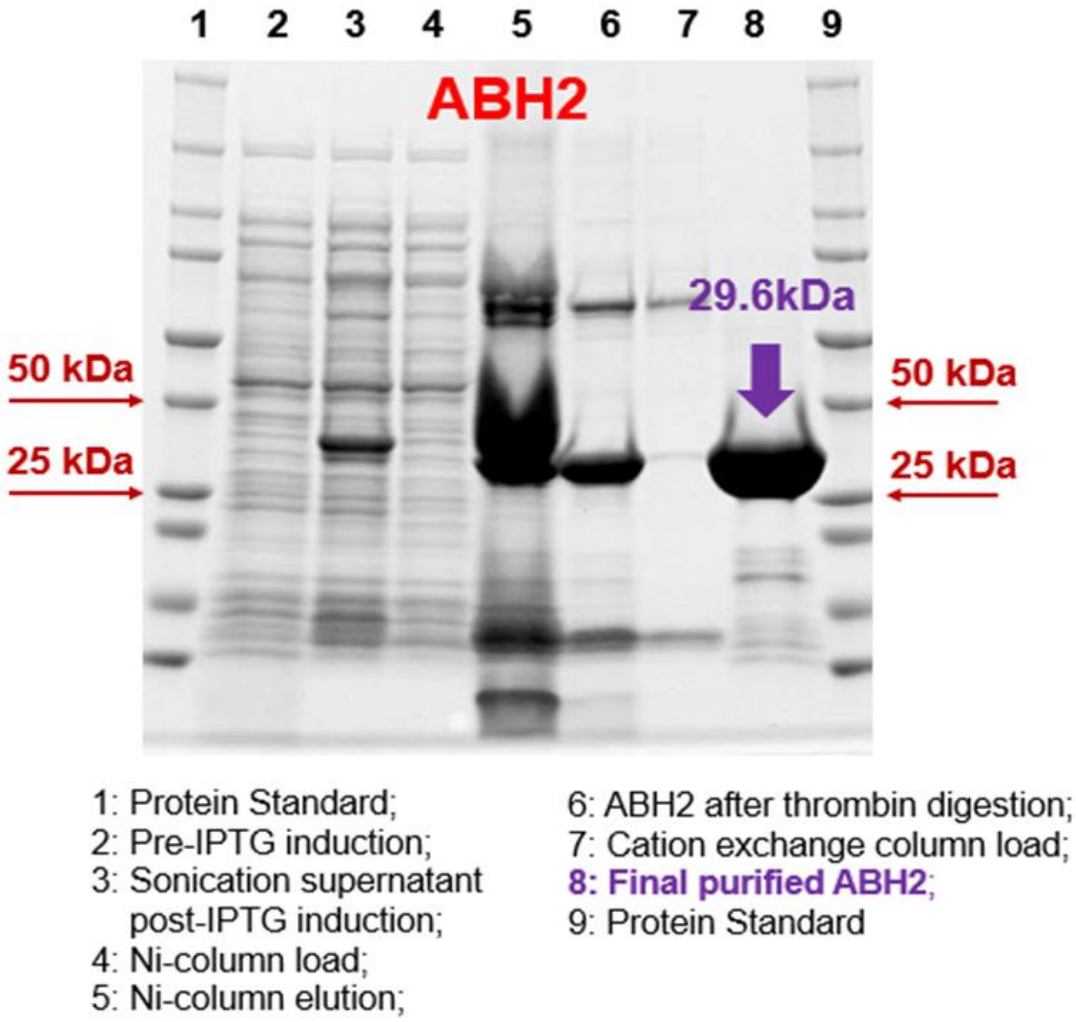

Figure S4. Polyacrylamide gel electrophoresis of purified ABH2 protein.

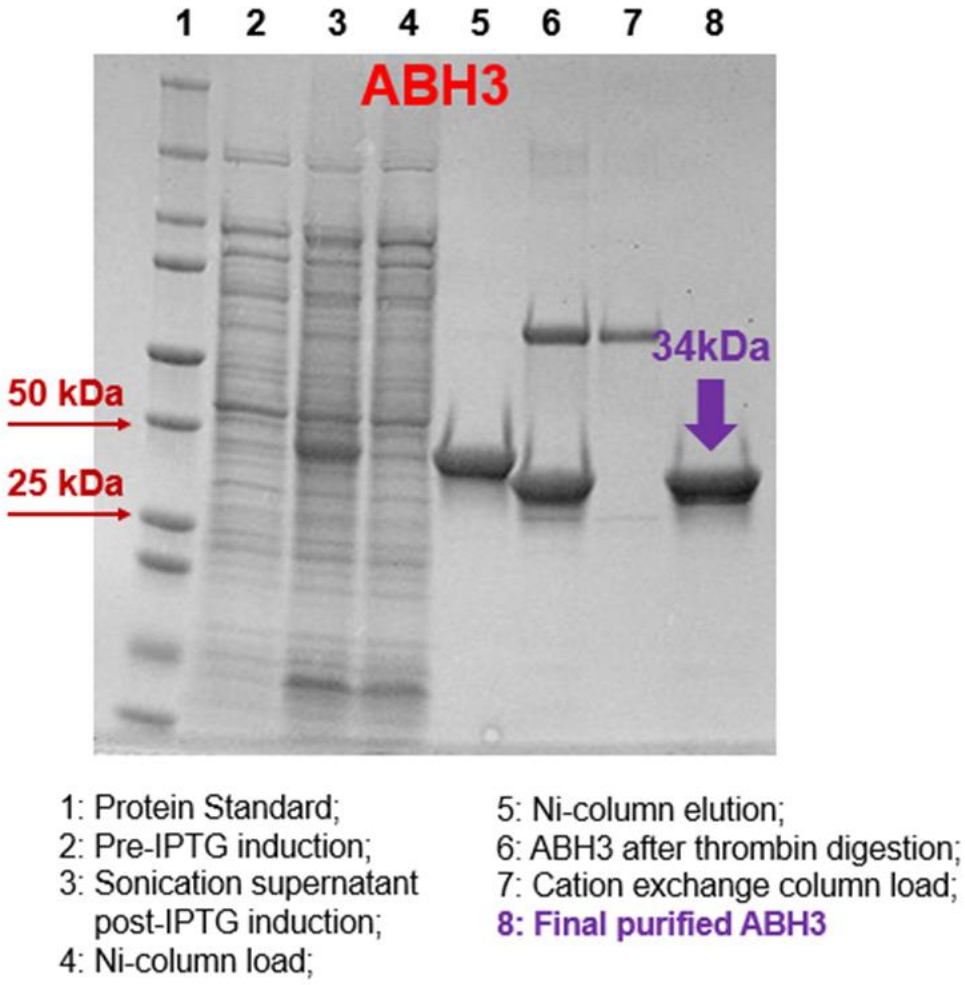

Figure S5. Polyacrylamide gel electrophoresis of purified ABH3 protein.

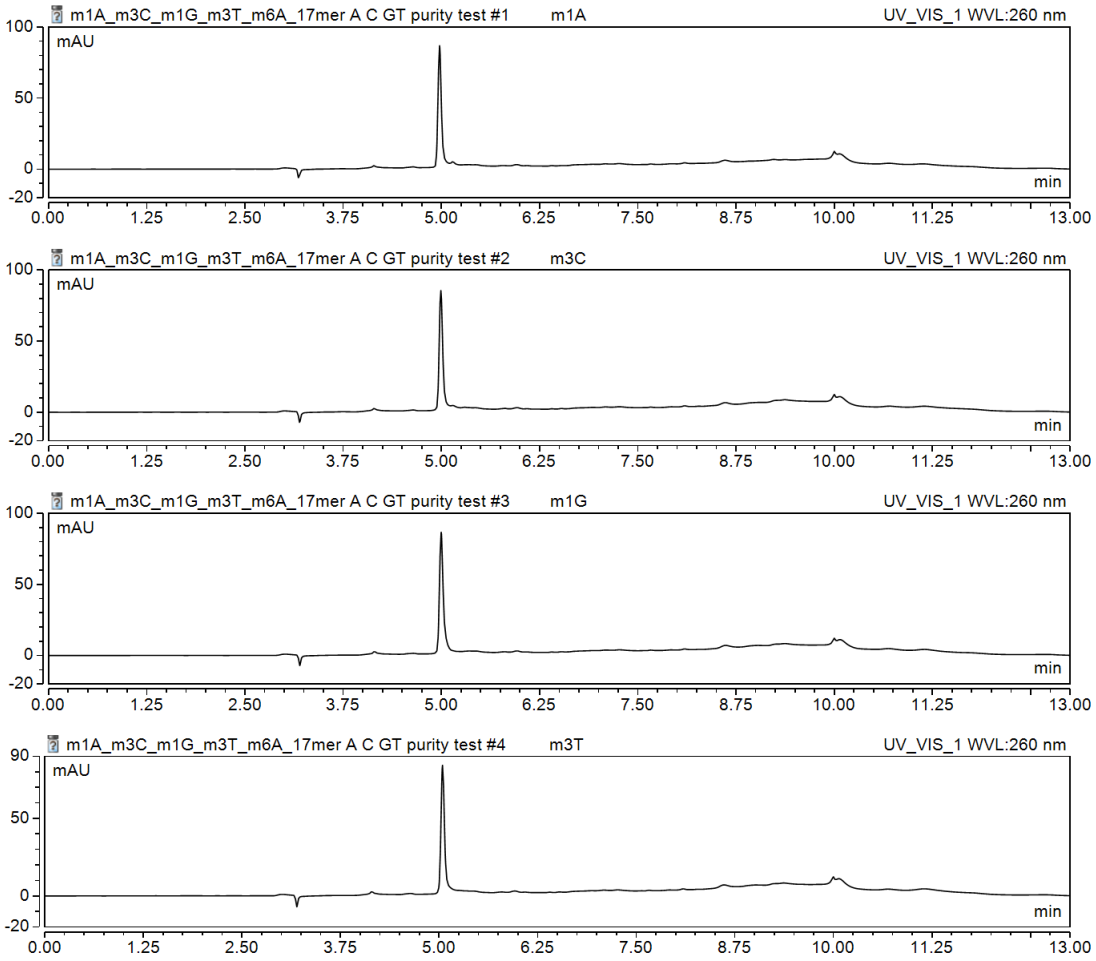

Figure S6. Purity test of oligonucleotides containing the four alkyl adducts by HPLC. The retention time of the oligos is around 5.0 min. For each test, 100pmol oligo was injected. From top to bottom, the sub-figures represent 16mer oligos containing m1A, m3C, m1G, and m3T.

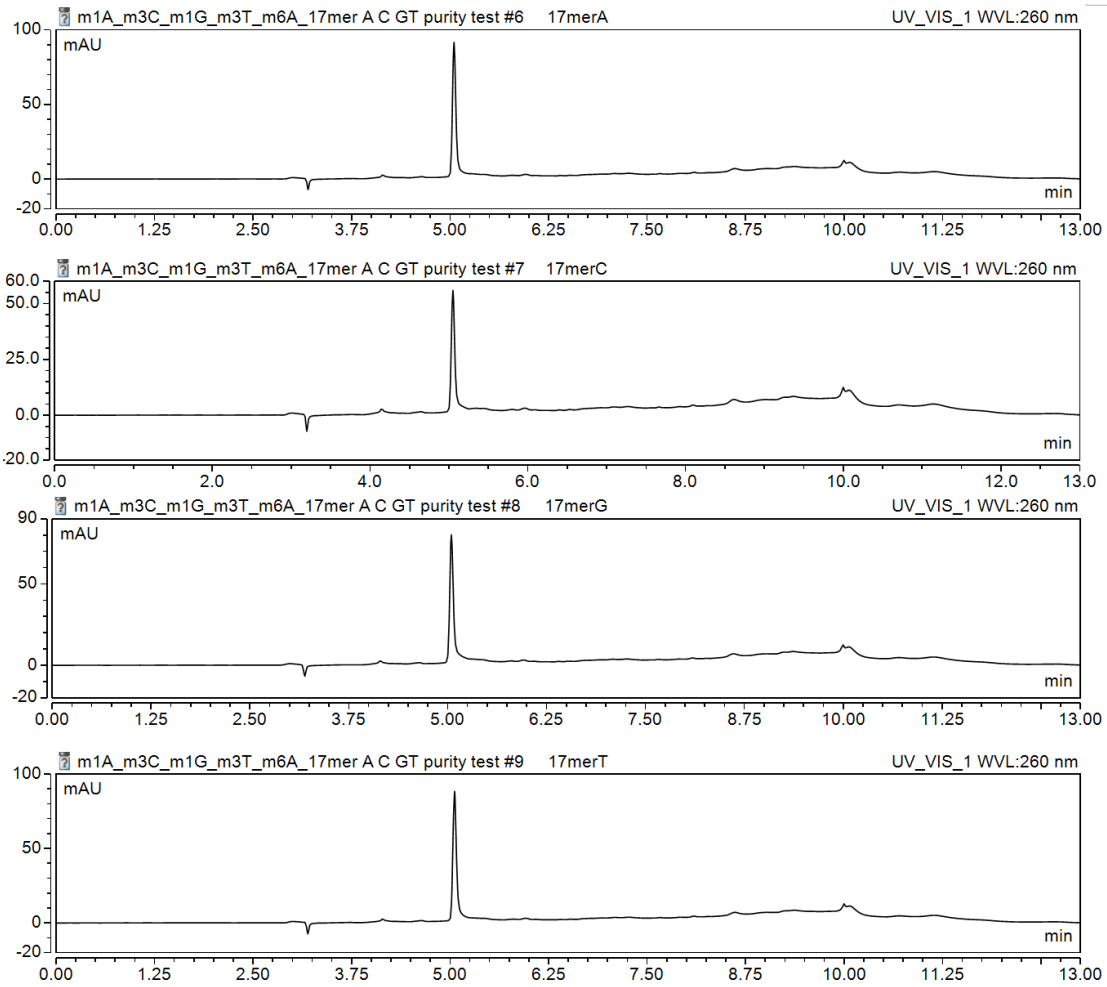

Figure S7. Purity test of oligonucleotides containing A/C/G/T in the 17mer complementary strands by HPLC. The retention time of the oligo is around 5.0 min. For each test, 100pmol oligo was injected. From top to bottom, the sub-figures represent 17mer oligos containing A, C, G, and T.

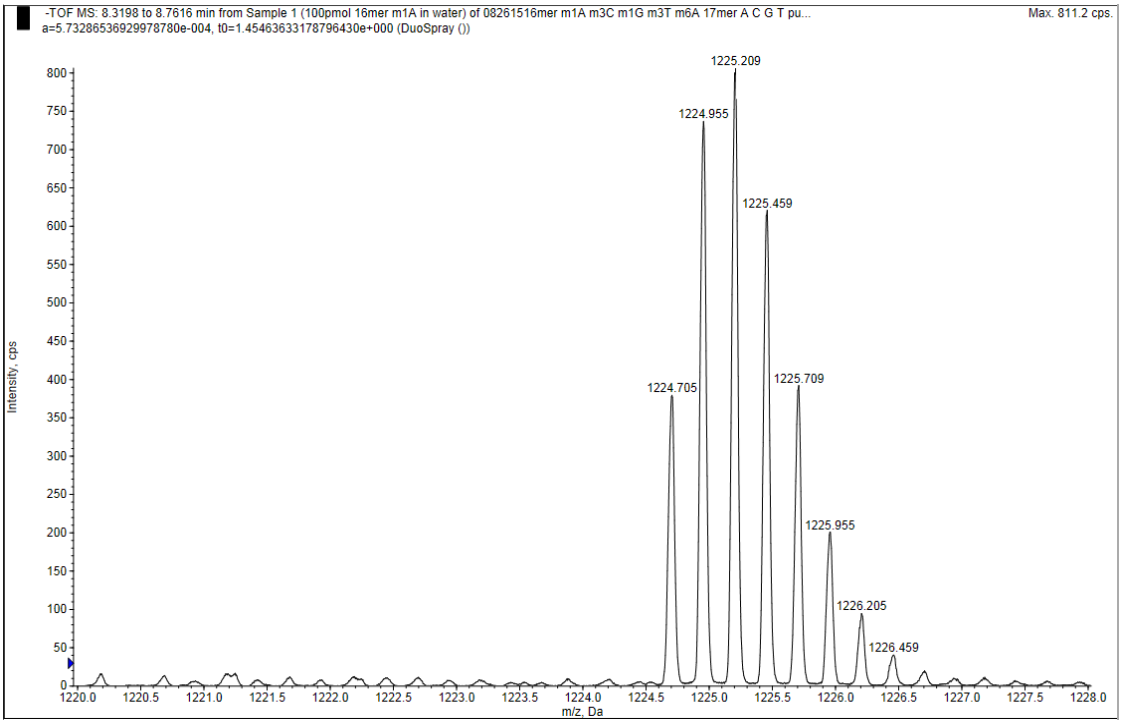

Figure S8. ESI-TOF analysis of 16mer oligonucleotide containing m1A.

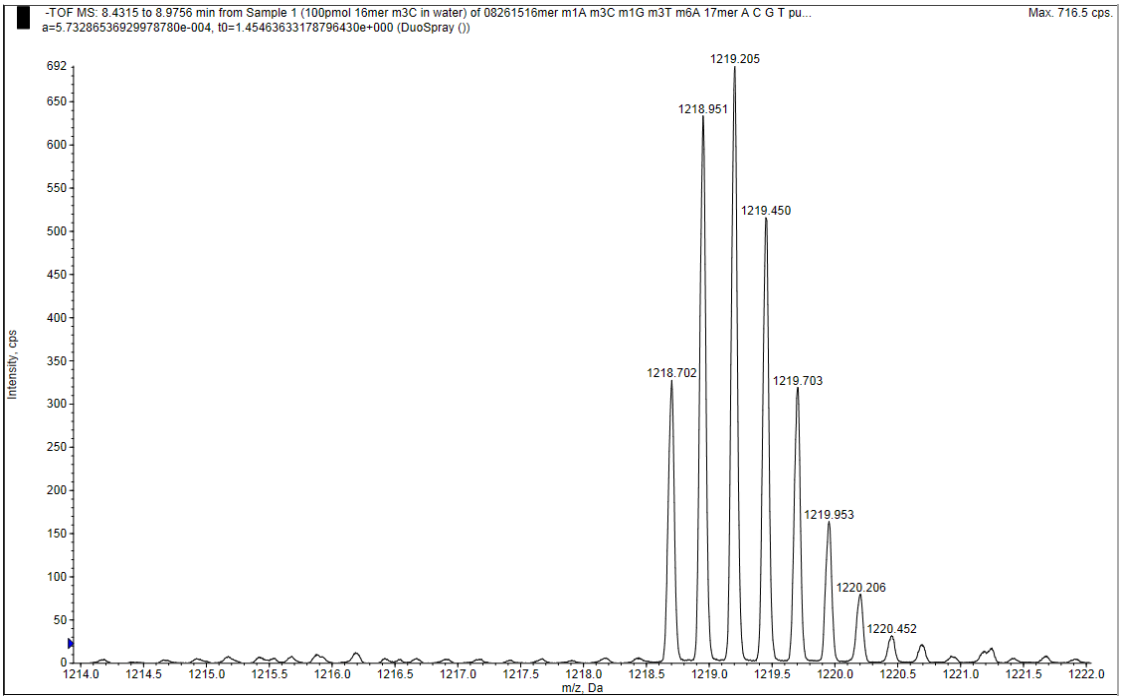

Figure S9. ESI-TOF analysis of 16mer oligonucleotide containing m3C.

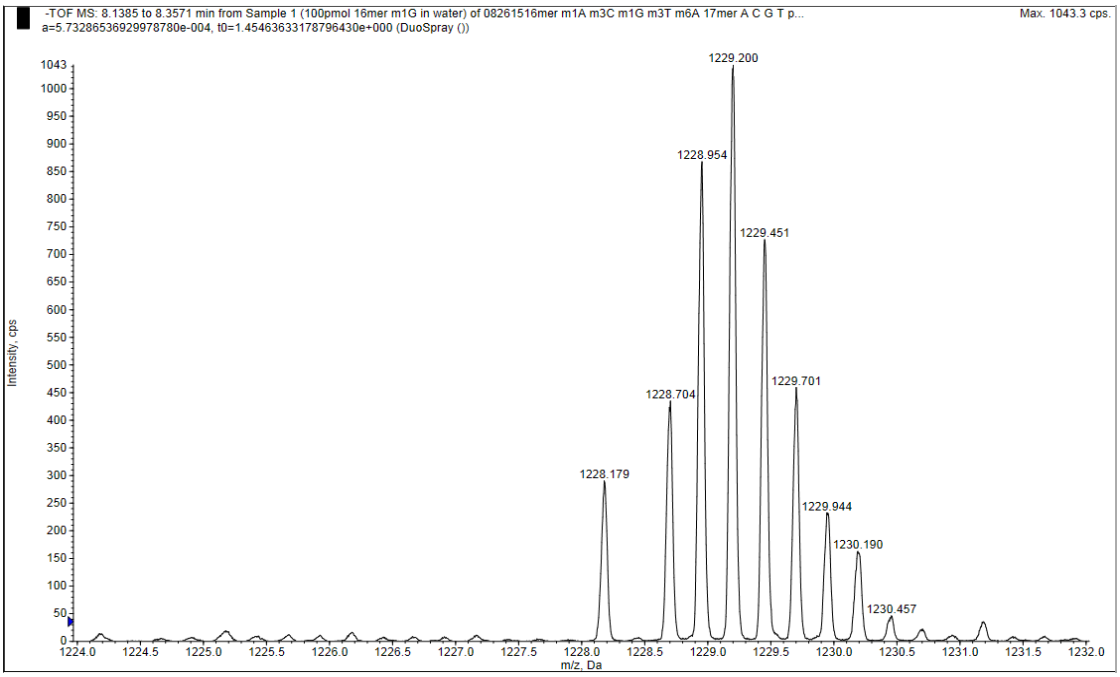

Figure S10. ESI-TOF analysis of 16mer oligonucleotide containing m1G. The peak at 1228.179 is from an impurity. The peak at 1228.704 is the monoisotopic peak of the 16mer containing m1G.

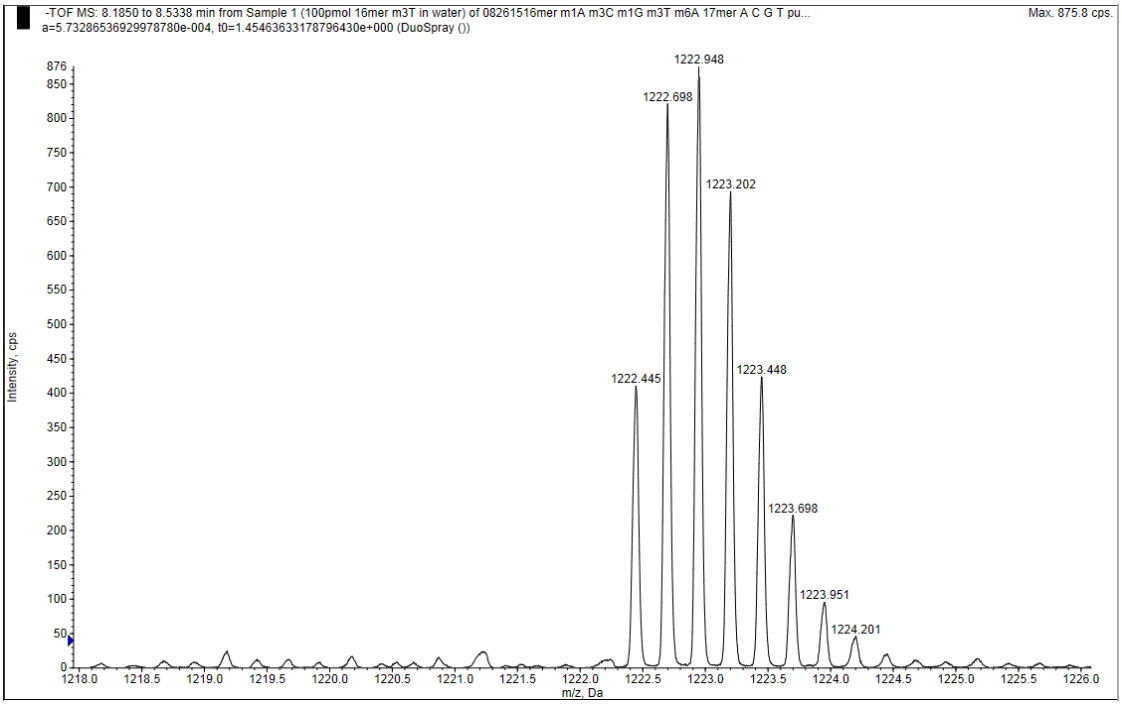

Figure S11. ESI-TOF analysis of 16mer oligonucleotide containing m3T.

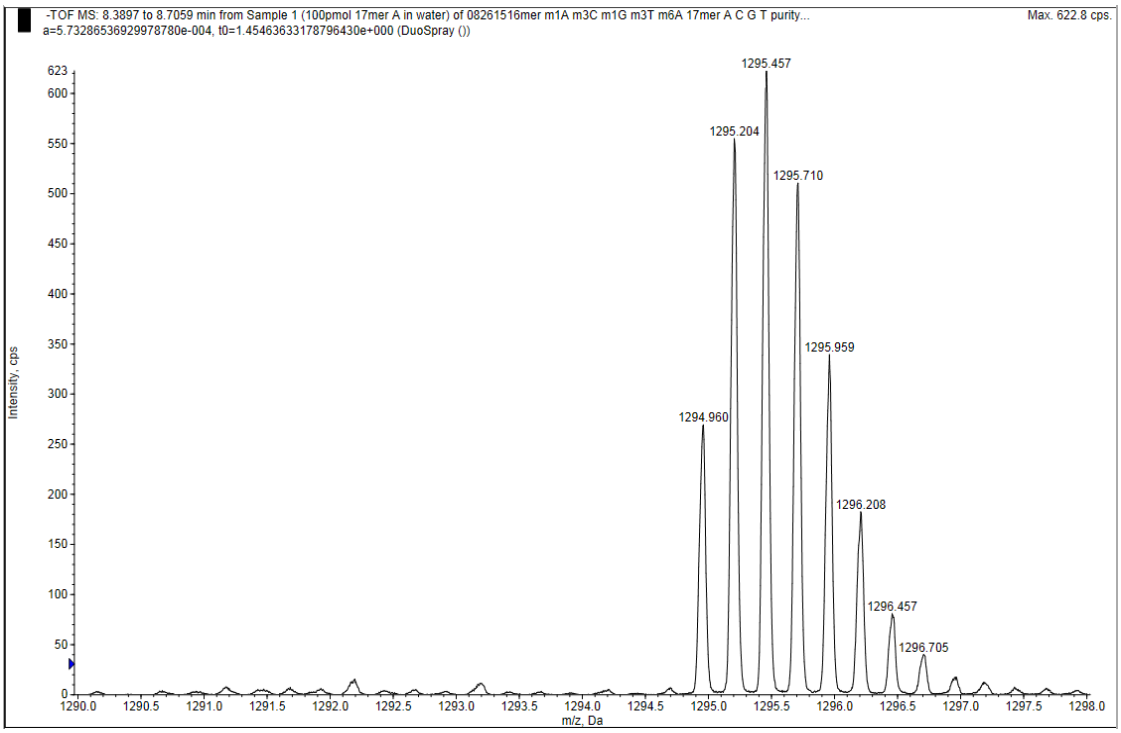

Figure S12. ESI-TOF analysis of 17mer oligonucleotide containing A.

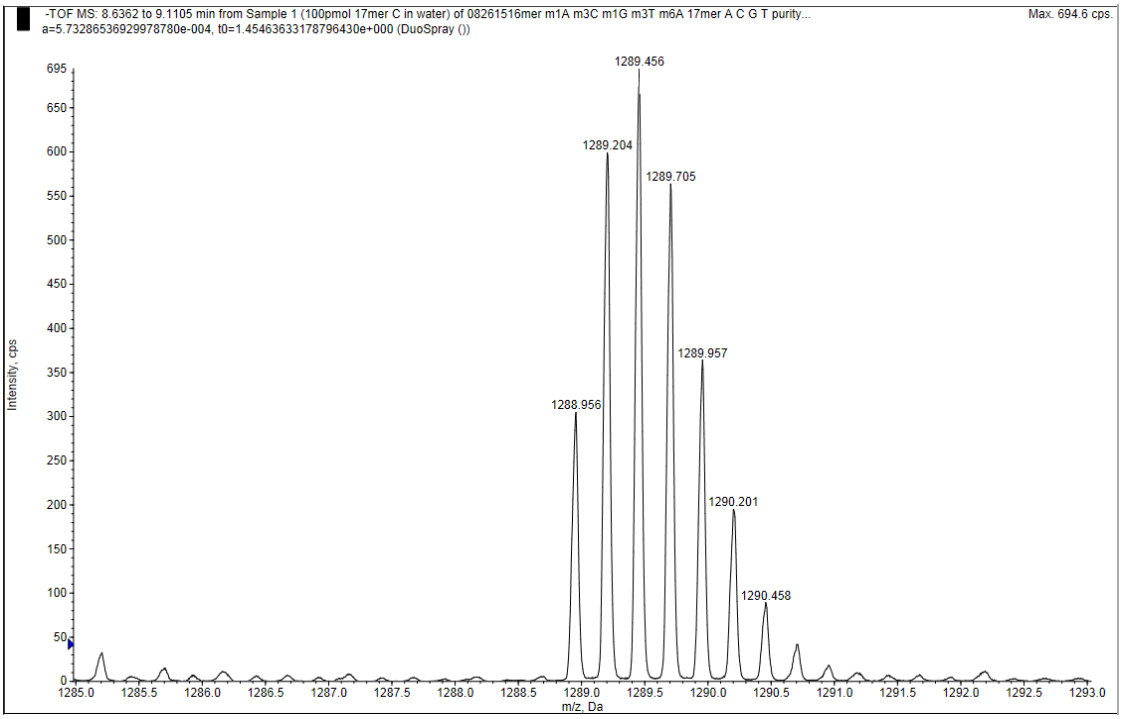

Figure S13. ESI-TOF analysis of 17mer oligonucleotide containing C.

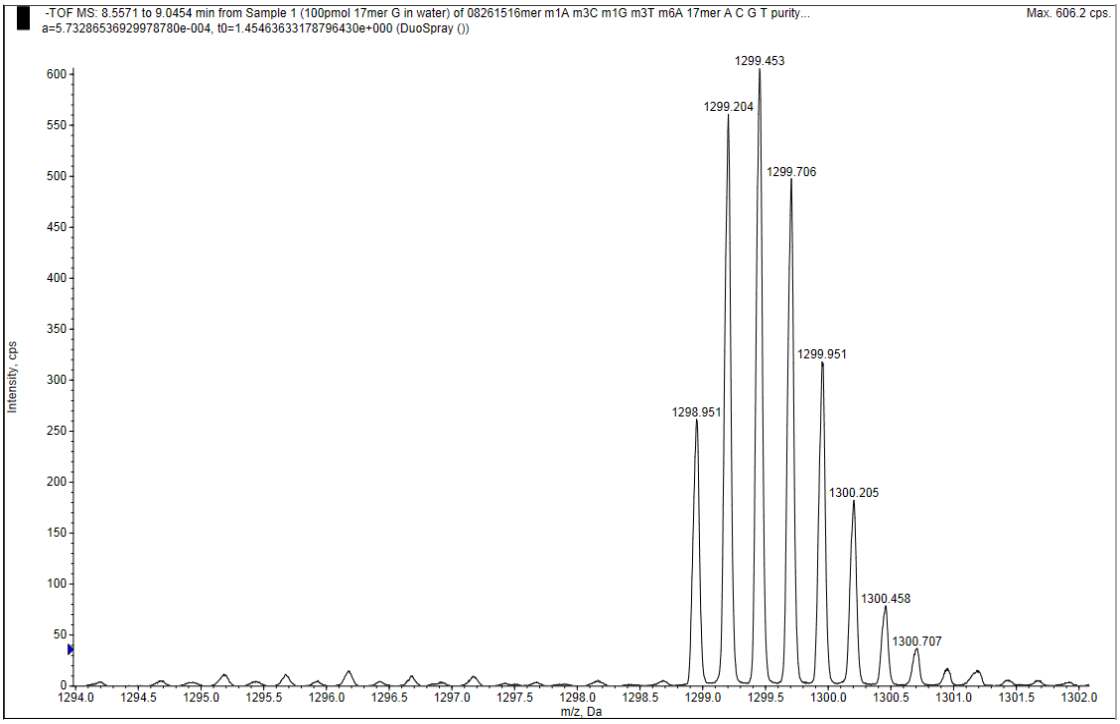

Figure S14. ESI-TOF analysis of 17mer oligonucleotide containing G.

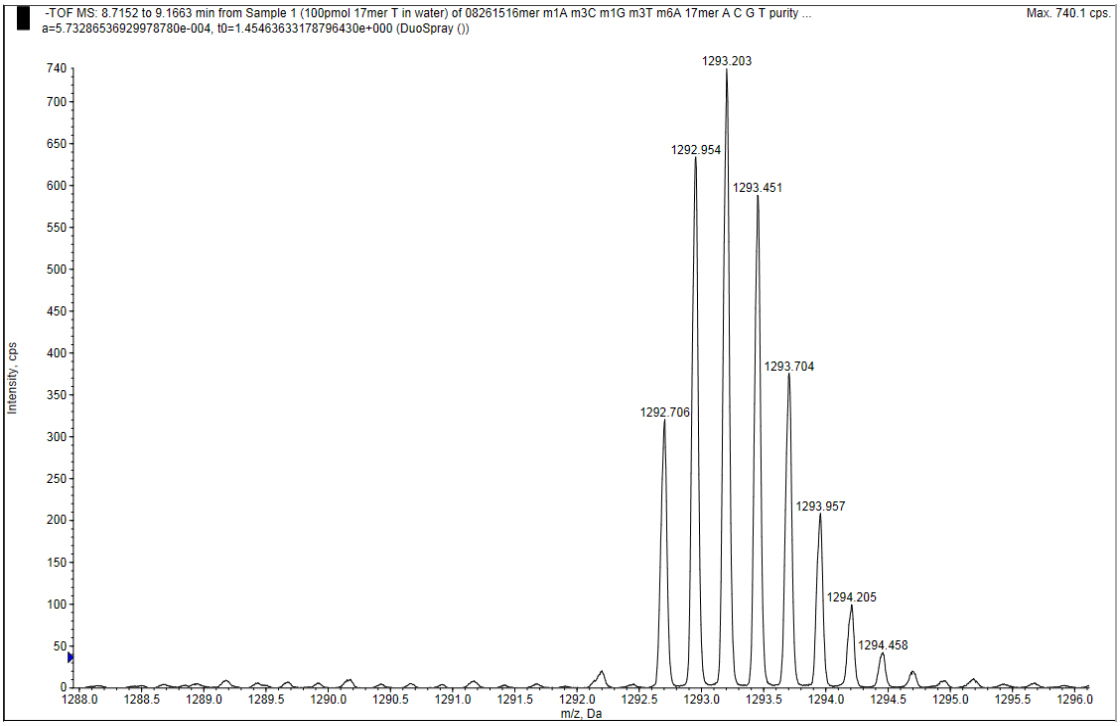

Figure S15. ESI-TOF analysis of 17mer oligonucleotide containing T.

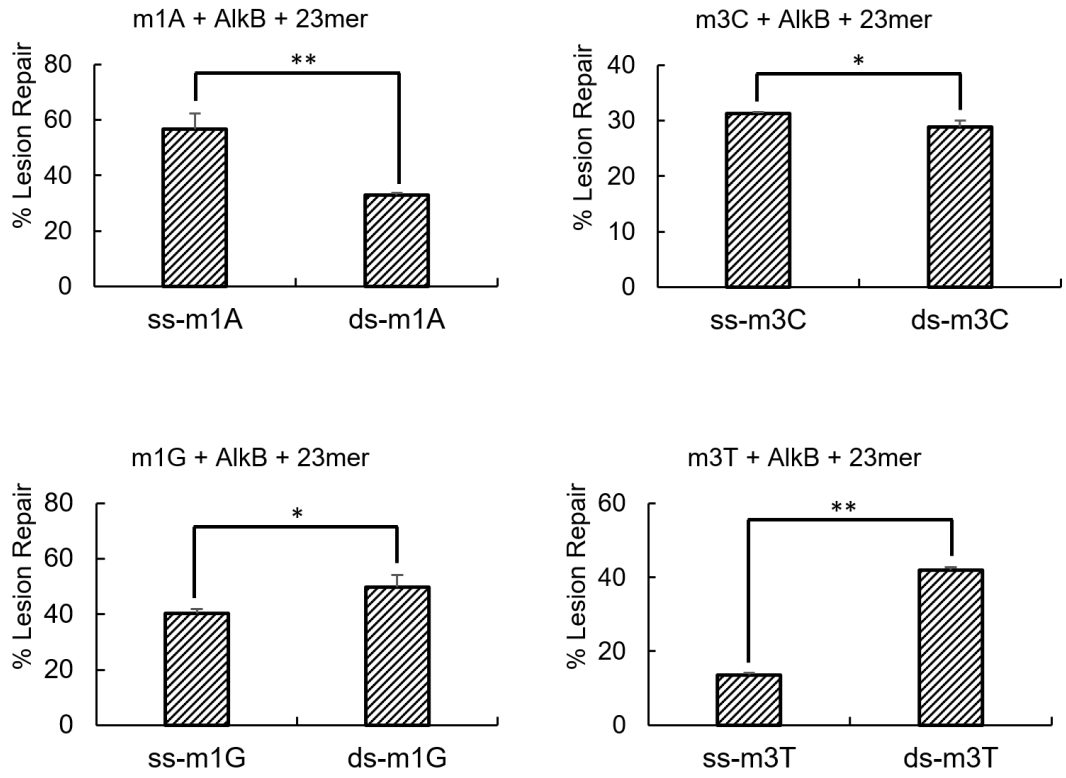

Figure S16. Repair efficiency of AlkB on different adducts in ss- and perfect match ds-DNA with non-mutagenic pairing and additional 15μM unrelated ss-23mer DNA. The error bars represent the standard deviation from triplicate experiments. The significance of the difference between ss-DNA and ds-DNA with non-mutagenic pairing was tested using the Student's two-tailed t test. * indicates p-value <0.05 and ** indicates p-value <0.01.

TABLES

Table 1. Initial velocity measurements of AlkB on the four DNA substrates in ss- and perfect match ds-DNA. Each reaction was carried out in triplicate. The significance of the difference between ss- and ds-DNA reactions was tested using the Student's two-tailed t test. The p-values are statistically significant.

	ss-DNA [$\mu M \cdot min^{-1}$]	ds-DNA [$\mu M \cdot min^{-1}$]	p-value
m1A	1.17 ± 0.05	1.03 ± 0.02	0.012
m3C	1.13 ± 0.01	0.84 ± 0.02	<0.001
m1G	0.57 ± 0.03	0.88 ± 0.02	<0.001
m3T	0.03 ± 0.01	0.10 ± 0.01	0.001

Table S1. Calculated and observed monoisotopic molecular weight and m/z value of oligonucleotides present in the enzyme repair reactions. The sequence of the 16mer oligos is 5'-GAAGACCTXGGCGTCC-3', where X indicates the position of the lesions and the repaired products. The sequence of the 17mer oligos is 5'-TGGACGCCYAGGTCTTC-3', where Y represents the position incorporating the complementary bases.

Lesion or base	MW (calculated) of neutral species	m/z (calculate) -4 charge peak	m/z (observed) -4 charge peak
16mer m1A	4902.88	1224.71	1224.71
16mer m3C	4878.87	1218.71	1218.70
16mer m1G	4918.87	1228.71	1228.70
16mer m3T	4893.87	1222.46	1222.45
16mer A	4888.86	1221.21	1221.20
16mer C	4864.85	1215.20	1215.20
16mer G	4904.86	1225.21	1225.20
16mer T	4879.85	1218.95	1218.96
17mer A	5183.90	1294.97	1294.96
17mer C	5159.88	1288.96	1288.96
17mer G	5199.89	1298.97	1298.95
17mer T	5174.88	1292.71	1292.71

Table S2. Repair efficiency of the three enzymes on four lesions (Figure 3). Each repair reaction was carried out in triplicate.

AlkB repair of indicated lesion in ss-DNA and ds-DNA.

Lesion	Repair ratio in ss-DNA and ds-DNA (%)				
	ss-DNA	lesion+A	lesion+C	lesion+G	lesion+T
m1A	53.4 ± 1.5	52.7 ± 4.0	39.0 ± 0.2	40.3 ± 1.1	21.8 ± 1.4
m3C	79.0 ± 1.2	81.4 ± 0.8	70.3 ± 1.7	67.1 ± 1.4	65.8 ± 1.4
m1G	48.2 ± 1.1	50.0 ± 0.3	69.2 ± 0.9	43.7 ± 1.8	55.5 ± 0.5
m3T	12.0 ± 0.4	21.0 ± 0.5	27.8 ± 0.3	12.5 ± 0.0	16.8 ± 0.5

ABH2 repair of indicated lesion in ss-DNA and ds-DNA.

Lesion	Repair ratio in ss-DNA and ds-DNA (%)				
	ss-DNA	lesion+A	lesion+C	lesion+G	lesion+T
m1A	64.5 ± 1.4	98.3 ± 0.2	96.7 ± 1.0	78.7 ± 1.5	77.9 ± 1.2
m3C	59.7 ± 0.6	93.5 ± 0.4	84.2 ± 0.5	88.8 ± 1.2	54.0 ± 0.3
m1G	22.2 ± 0.7	70.9 ± 0.6	62.3 ± 1.2	44.7 ± 0.7	70.7 ± 1.5
m3T	15.9 ± 1.2	94.1 ± 0.9	92.5 ± 0.7	93.5 ± 0.4	93.3 ± 0.8

ABH3 repair of indicated lesion in ss-DNA and ds-DNA.

Lesion	Repair ratio in ss-DNA and ds-DNA (%)				
	ss-DNA	lesion+A	lesion+C	lesion+G	lesion+T
m1A	72.0 ± 1.5	13.0 ± 0.7	51.3 ± 2.2	9.8 ± 0.8	9.1 ± 0.5
m3C	57.0 ± 0.4	13.6 ± 0.2	45.2 ± 1.1	10.5 ± 0.5	20.9 ± 0.2

Table S3. Enzyme concentrations of the three enzymes in the repair reactions. For each repair reaction, 5 μM 16mer DNA oligo and the corresponding concentration of the enzyme in this table were used. * the repair efficiencies of ABH3 on m1G and m3T were extremely low and the products were undetectable even with addition of large excess amount of the enzyme, such as adding 5μM enzyme to 5μM DNA adducts.

lesion	AlkB (μM)	ABH2 (μM)	ABH3 (μM)
m1A	0.25	0.31	0.63
m3C	0.08	0.31	0.40
m1G	0.75	5.00	-*
m3T	1.50	1.25	-*

Table S4. Enzyme concentrations of AlkB in the reactions of measuring the initial velocity. For each repair reaction, 5 μM 16mer DNA oligo and the corresponding concentration of AlkB in this table were used.

Lesion	AlkB (μM)
m1A	0.1
m3C	0.035
m1G	2.0
m3T	1.5

Table S5. Repair efficiency of the AlkB enzyme on four lesions with additional 15μM unrelated 23mer DNA. The ds-DNA reactions were carried out with the perfect match base pairs of each adduct. The corresponding concentrations of AlkB were listed in the table. Each repair reaction was carried out in triplicate.

lesion	ss-DNA (%)	ds-DNA (%)	p-value	AlkB (μM)
m1A	56.8 ± 5.5	33.0 ± 0.7	0.016	0.50
m3C	31.3 ± 0.2	28.8 ± 1.3	0.028	0.15
m1G	40.3 ± 1.6	49.8 ± 4.4	0.024	2.25
m3T	13.7± 0.4	41.9 ± 0.8	<0.001	4.50

ACKNOWLEDGMENTS

The authors want to thank the RI-INBRE program, its directors Prof. Zahir Shaikh and Prof. David Rowley, and the staff Dr. Al Bach, Ms. Kim Andrews and Ms. Patricia Murray for their kind help. We also want to thank Prof. Bingfang Yan, Prof. Bongsup Cho, Prof. Gongqin Sun, Prof. Fatemeh Akhlaghi, Dr. James C. Delaney, Dr. Lifang Xu, Ms. Sravani Adusumalli, Mr. Aram Babcock, Mr. Ang Cai, Ms. Yixin Cui, and Ms. Zhengxi Wei for their support and helpful discussions. This work was supported by an Institutional Development Award from the National Institute of General Medical Sciences of the National Institutes of Health under grant number 2 P20 GM103430, and a Medical Research Funds grant from the Rhode Island foundation (to D. L.). This work was also supported, in whole or in part, by National Institutes of Health Grants P01 CA26731, R37 CA080024, and P30 ES002109 (to J. M. E.). C. L. D. is a Howard Hughes Medical Institute Investigator.

REFERENCES

(1) Yi, C.; He, C. DNA Repair by Reversal of DNA Damage. *Cold Spring Harb. Perspect. Biol.* **2013**, *5* (1), a012575. https://doi.org/10.1101/cshperspect.a012575.

(2) Fu, D.; Samson, L. D. Direct Repair of 3,N(4)-Ethenocytosine by the Human ALKBH2 Dioxygenase Is Blocked by the AAG/MPG Glycosylase. *DNA Repair (Amst.)* **2012**, *11* (1), 46–52. https://doi.org/10.1016/j.dnarep.2011.10.004.

(3) Shrivastav, N.; Li, D.; Essigmann, J. M. Chemical Biology of Mutagenesis and DNA Repair: Cellular Responses to DNA Alkylation. *Carcinogenesis* **2010**, *31* (1), 59–70. https://doi.org/10.1093/carcin/bgp262.

(4) Aravind, L.; Koonin, E. V. The DNA-Repair Protein AlkB, EGL-9, and Leprecan Define New Families of 2-Oxoglutarate- and Iron-Dependent Dioxygenases. *Genome Biol.* **2001**, *2* (3), RESEARCH0007.

(5) Sedgwick, B.; Bates, P. A.; Paik, J.; Jacobs, S. C.; Lindahl, T. Repair of Alkylated DNA: Recent Advances. *DNA Repair (Amst.)* **2007**, *6* (4), 429–442. https://doi.org/10.1016/j.dnarep.2006.10.005.

(6) Silvestrov, P.; Müller, T. A.; Clark, K. N.; Hausinger, R. P.; Cisneros, G. A. Homology Modeling, Molecular Dynamics, and Site-Directed Mutagenesis Study of AlkB Human Homolog 1 (ALKBH1). *J. Mol. Graph. Model.* **2014**, *54*, 123–130. https://doi.org/10.1016/j.jmgm.2014.10.013.

(7) Westbye, M. P.; Feyzi, E.; Aas, P. A.; Vågbø, C. B.; Talstad, V. A.; Kavli, B.; Hagen, L.; Sundheim, O.; Akbari, M.; Liabakk, N.-B.; et al. Human AlkB Homolog 1 Is a Mitochondrial Protein That Demethylates 3-Methylcytosine in

DNA and RNA. *J. Biol. Chem.* **2008**, *283* (36), 25046–25056. https://doi.org/10.1074/jbc.M803776200.

(8) Li, M.-M.; Nilsen, A.; Shi, Y.; Fusser, M.; Ding, Y.-H.; Fu, Y.; Liu, B.; Niu, Y.; Wu, Y.-S.; Huang, C.-M.; et al. ALKBH4-Dependent Demethylation of Actin Regulates Actomyosin Dynamics. *Nat Commun* **2013**, *4*, 1832. https://doi.org/10.1038/ncomms2863.

(9) Bjørnstad, L. G.; Zoppellaro, G.; Tomter, A. B.; Falnes, P. Ø.; Andersson, K. K. Spectroscopic and Magnetic Studies of Wild-Type and Mutant Forms of the Fe(II)- and 2-Oxoglutarate-Dependent Decarboxylase ALKBH4. *Biochem. J.* **2011**, *434* (3), 391–398. https://doi.org/10.1042/BJ20101667.

(10) Fu, D.; Jordan, J. J.; Samson, L. D. Human ALKBH7 Is Required for Alkylation and Oxidation-Induced Programmed Necrosis. *Genes Dev.* **2013**, *27* (10), 1089–1100. https://doi.org/10.1101/gad.215533.113.

(11) Wang, G.; He, Q.; Feng, C.; Liu, Y.; Deng, Z.; Qi, X.; Wu, W.; Mei, P.; Chen, Z. The Atomic Resolution Structure of Human AlkB Homolog 7 (ALKBH7), a Key Protein for Programmed Necrosis and Fat Metabolism. *J. Biol. Chem.* **2014**, *289* (40), 27924–27936. https://doi.org/10.1074/jbc.M114.590505.

(12) Fu, D.; Brophy, J. A. N.; Chan, C. T. Y.; Atmore, K. A.; Begley, U.; Paules, R. S.; Dedon, P. C.; Begley, T. J.; Samson, L. D. Human AlkB Homolog ABH8 Is a TRNA Methyltransferase Required for Wobble Uridine Modification and DNA Damage Survival. *Mol. Cell. Biol.* **2010**, *30* (10), 2449–2459. https://doi.org/10.1128/MCB.01604-09.

(13) Fu, Y.; Dai, Q.; Zhang, W.; Ren, J.; Pan, T.; He, C. The AlkB Domain of Mammalian ABH8 Catalyzes Hydroxylation of 5-Methoxycarbonylmethyluridine at the Wobble Position of TRNA. *Angew. Chem. Int. Ed. Engl.* **2010**, *49* (47), 8885–8888. https://doi.org/10.1002/anie.201001242.

(14) van den Born, E.; Vågbø, C. B.; Songe-Møller, L.; Leihne, V.; Lien, G. F.; Leszczynska, G.; Malkiewicz, A.; Krokan, H. E.; Kirpekar, F.; Klungland, A.; et al. ALKBH8-Mediated Formation of a Novel Diastereomeric Pair of Wobble Nucleosides in Mammalian TRNA. *Nat Commun* **2011**, *2*, 172. https://doi.org/10.1038/ncomms1173.

(15) Pastore, C.; Topalidou, I.; Forouhar, F.; Yan, A. C.; Levy, M.; Hunt, J. F. Crystal Structure and RNA Binding Properties of the RNA Recognition Motif (RRM) and AlkB Domains in Human AlkB Homolog 8 (ABH8), an Enzyme Catalyzing TRNA Hypermodification. *J. Biol. Chem.* **2012**, *287* (3), 2130–2143. https://doi.org/10.1074/jbc.M111.286187.

(16) Zdżalik, D.; Vågbø, C. B.; Kirpekar, F.; Davydova, E.; Puścian, A.; Maciejewska, A. M.; Krokan, H. E.; Klungland, A.; Tudek, B.; van den Born, E.; et al. Protozoan ALKBH8 Oxygenases Display Both DNA Repair and TRNA Modification Activities. *PLoS ONE* **2014**, *9* (6), e98729. https://doi.org/10.1371/journal.pone.0098729.

(17) Xu, C.; Liu, K.; Tempel, W.; Demetriades, M.; Aik, W.; Schofield, C. J.; Min, J. Structures of Human ALKBH5 Demethylase Reveal a Unique Binding Mode

for Specific Single-Stranded N6-Methyladenosine RNA Demethylation. *J. Biol. Chem.* **2014**, *289* (25), 17299–17311. https://doi.org/10.1074/jbc.M114.550350.

(18) Shen, F.; Huang, W.; Huang, J.-T.; Xiong, J.; Yang, Y.; Wu, K.; Jia, G.-F.; Chen, J.; Feng, Y.-Q.; Yuan, B.-F.; et al. Decreased N(6)-Methyladenosine in Peripheral Blood RNA from Diabetic Patients Is Associated with FTO Expression Rather than ALKBH5. *J. Clin. Endocrinol. Metab.* **2015**, *100* (1), E148-154. https://doi.org/10.1210/jc.2014-1893.

(19) Thalhammer, A.; Bencokova, Z.; Poole, R.; Loenarz, C.; Adam, J.; O'Flaherty, L.; Schödel, J.; Mole, D.; Giaslakiotis, K.; Schofield, C. J.; et al. Human AlkB Homologue 5 Is a Nuclear 2-Oxoglutarate Dependent Oxygenase and a Direct Target of Hypoxia-Inducible Factor 1α (HIF-1α). *PLoS ONE* **2011**, *6* (1), e16210. https://doi.org/10.1371/journal.pone.0016210.

(20) Zheng, G.; Dahl, J. A.; Niu, Y.; Fedorcsak, P.; Huang, C.-M.; Li, C. J.; Vågbø, C. B.; Shi, Y.; Wang, W.-L.; Song, S.-H.; et al. ALKBH5 Is a Mammalian RNA Demethylase That Impacts RNA Metabolism and Mouse Fertility. *Mol. Cell* **2013**, *49* (1), 18–29. https://doi.org/10.1016/j.molcel.2012.10.015.

(21) Feng, C.; Liu, Y.; Wang, G.; Deng, Z.; Zhang, Q.; Wu, W.; Tong, Y.; Cheng, C.; Chen, Z. Crystal Structures of the Human RNA Demethylase Alkbh5 Reveal Basis for Substrate Recognition. *J. Biol. Chem.* **2014**, *289* (17), 11571–11583. https://doi.org/10.1074/jbc.M113.546168.

(22) Meyer, K. D.; Jaffrey, S. R. The Dynamic Epitranscriptome: N6-Methyladenosine and Gene Expression Control. *Nat. Rev. Mol. Cell Biol.* **2014**, *15* (5), 313–326. https://doi.org/10.1038/nrm3785.

(23) Liu, C.; Mou, S.; Cai, Y. FTO Gene Variant and Risk of Overweight and Obesity among Children and Adolescents: A Systematic Review and Meta-Analysis. *PLoS ONE* **2013**, *8* (11), e82133. https://doi.org/10.1371/journal.pone.0082133.

(24) Zhao, X.; Yang, Y.; Sun, B.-F.; Zhao, Y.-L.; Yang, Y.-G. FTO and Obesity: Mechanisms of Association. *Curr. Diab. Rep.* **2014**, *14* (5), 486. https://doi.org/10.1007/s11892-014-0486-0.

(25) Chandola, U.; Das, R.; Panda, B. Role of the N6-Methyladenosine RNA Mark in Gene Regulation and Its Implications on Development and Disease. *Brief Funct Genomics* **2014**, *14* (3), 169–179. https://doi.org/10.1093/bfgp/elu039.

(26) Li, P.; Gao, S.; Wang, L.; Yu, F.; Li, J.; Wang, C.; Li, J.; Wong, J. ABH2 Couples Regulation of Ribosomal DNA Transcription with DNA Alkylation Repair. *Cell Rep* **2013**, *4* (4), 817–829. https://doi.org/10.1016/j.celrep.2013.07.027.

(27) Chen, B.; Liu, H.; Sun, X.; Yang, C.-G. Mechanistic Insight into the Recognition of Single-Stranded and Double-Stranded DNA Substrates by ABH2 and ABH3. *Mol Biosyst* **2010**, *6* (11), 2143–2149. https://doi.org/10.1039/c005148a.

(28) Yang, C.-G.; Yi, C.; Duguid, E. M.; Sullivan, C. T.; Jian, X.; Rice, P. A.; He, C. Crystal Structures of DNA/RNA Repair Enzymes AlkB and ABH2 Bound to DsDNA. *Nature* **2008**, *452* (7190), 961–965. https://doi.org/10.1038/nature06889.

(29) Yi, C.; Jia, G.; Hou, G.; Dai, Q.; Zhang, W.; Zheng, G.; Jian, X.; Yang, C.-G.; Cui, Q.; He, C. Iron-Catalysed Oxidation Intermediates Captured in a DNA Repair Dioxygenase. *Nature* **2010**, *468* (7321), 330–333. https://doi.org/10.1038/nature09497.

(30) Nay, S. L.; Lee, D.-H.; Bates, S. E.; O'Connor, T. R. Alkbh2 Protects against Lethality and Mutation in Primary Mouse Embryonic Fibroblasts. *DNA Repair (Amst.)* **2012**, *11* (5), 502–510. https://doi.org/10.1016/j.dnarep.2012.02.005.

(31) Dango, S.; Mosammaparast, N.; Sowa, M. E.; Xiong, L.-J.; Wu, F.; Park, K.; Rubin, M.; Gygi, S.; Harper, J. W.; Shi, Y. DNA Unwinding by ASCC3 Helicase Is Coupled to ALKBH3-Dependent DNA Alkylation Repair and Cancer Cell Proliferation. *Mol. Cell* **2011**, *44* (3), 373–384. https://doi.org/10.1016/j.molcel.2011.08.039.

(32) Koike, K.; Ueda, Y.; Hase, H.; Kitae, K.; Fusamae, Y.; Masai, S.; Inagaki, T.; Saigo, Y.; Hirasawa, S.; Nakajima, K.; et al. Anti-Tumor Effect of AlkB Homolog 3 Knockdown in Hormone- Independent Prostate Cancer Cells. *Curr Cancer Drug Targets* **2012**, *12* (7), 847–856.

(33) Calvo, J. A.; Meira, L. B.; Lee, C.-Y. I.; Moroski-Erkul, C. A.; Abolhassani, N.; Taghizadeh, K.; Eichinger, L. W.; Muthupalani, S.; Nordstrand, L. M.; Klungland, A.; et al. DNA Repair Is Indispensable for Survival after Acute Inflammation. *J. Clin. Invest.* **2012**, *122* (7), 2680–2689. https://doi.org/10.1172/JCI63338.

(34) Sundheim, O.; Vågbø, C. B.; Bjørås, M.; Sousa, M. M. L.; Talstad, V.; Aas, P. A.; Drabløs, F.; Krokan, H. E.; Tainer, J. A.; Slupphaug, G. Human ABH3

Structure and Key Residues for Oxidative Demethylation to Reverse DNA/RNA Damage. *EMBO J.* **2006**, *25* (14), 3389–3397. https://doi.org/10.1038/sj.emboj.7601219.

(35) Maciejewska, A. M.; Poznanski, J.; Kaczmarska, Z.; Krowisz, B.; Nieminuszczy, J.; Polkowska-Nowakowska, A.; Grzesiuk, E.; Kusmierek, J. T. AlkB Dioxygenase Preferentially Repairs Protonated Substrates: Specificity against Exocyclic Adducts and Molecular Mechanism of Action. *J. Biol. Chem.* **2013**, *288* (1), 432–441. https://doi.org/10.1074/jbc.M112.353342.

(36) Delaney, J. C.; Essigmann, J. M. Mutagenesis, Genotoxicity, and Repair of 1-Methyladenine, 3-Alkylcytosines, 1-Methylguanine, and 3-Methylthymine in AlkB Escherichia Coli. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101* (39), 14051–14056. https://doi.org/10.1073/pnas.0403489101.

(37) Delaney, J. C.; Smeester, L.; Wong, C.; Frick, L. E.; Taghizadeh, K.; Wishnok, J. S.; Drennan, C. L.; Samson, L. D.; Essigmann, J. M. AlkB Reverses Etheno DNA Lesions Caused by Lipid Oxidation in Vitro and in Vivo. *Nat. Struct. Mol. Biol.* **2005**, *12* (10), 855–860. https://doi.org/10.1038/nsmb996.

(38) Frick, L. E.; Delaney, J. C.; Wong, C.; Drennan, C. L.; Essigmann, J. M. Alleviation of 1,N6-Ethanoadenine Genotoxicity by the Escherichia Coli Adaptive Response Protein AlkB. *Proc. Natl. Acad. Sci. U.S.A.* **2007**, *104* (3), 755–760. https://doi.org/10.1073/pnas.0607377104.

(39) Li, D.; Delaney, J. C.; Page, C. M.; Chen, A. S.; Wong, C.; Drennan, C. L.; Essigmann, J. M. Repair of DNA Alkylation Damage by the Escherichia Coli

Adaptive Response Protein AlkB as Studied by ESI-TOF Mass Spectrometry. *J. Nucleic Acids* **2010**, *2010*, 369434. https://doi.org/10.4061/2010/369434.

(40) Li, D.; Delaney, J. C.; Page, C. M.; Yang, X.; Chen, A. S.; Wong, C.; Drennan, C. L.; Essigmann, J. M. Exocyclic Carbons Adjacent to the N6 of Adenine Are Targets for Oxidation by the Escherichia Coli Adaptive Response Protein AlkB. *J. Am. Chem. Soc.* **2012**, *134* (21), 8896–8901. https://doi.org/10.1021/ja3010094.

(41) Li, D.; Fedeles, B. I.; Shrivastav, N.; Delaney, J. C.; Yang, X.; Wong, C.; Drennan, C. L.; Essigmann, J. M. Removal of N-Alkyl Modifications from N(2)-Alkylguanine and N(4)-Alkylcytosine in DNA by the Adaptive Response Protein AlkB. *Chem. Res. Toxicol.* **2013**, *26* (8), 1182–1187. https://doi.org/10.1021/tx400096m.

(42) Singh, V.; Fedeles, B. I.; Li, D.; Delaney, J. C.; Kozekov, I. D.; Kozekova, A.; Marnett, L. J.; Rizzo, C. J.; Essigmann, J. M. Mechanism of Repair of Acrolein- and Malondialdehyde-Derived Exocyclic Guanine Adducts by the α-Ketoglutarate/Fe(II) Dioxygenase AlkB. *Chem. Res. Toxicol.* **2014**, *27* (9), 1619–1631. https://doi.org/10.1021/tx5002817.

(43) Chang, S.; Fedeles, B. I.; Wu, J.; Delaney, J. C.; Li, D.; Zhao, L.; Christov, P. P.; Yau, E.; Singh, V.; Jost, M.; et al. Next-Generation Sequencing Reveals the Biological Significance of the N(2),3-Ethenoguanine Lesion in Vivo. *Nucleic Acids Res.* **2015**, *43* (11), 5489–5500. https://doi.org/10.1093/nar/gkv243.

(44) Zheng, G.; Fu, Y.; He, C. Nucleic Acid Oxidation in DNA Damage Repair and Epigenetics. *Chem. Rev.* **2014**, *114* (8), 4602–4620. https://doi.org/10.1021/cr400432d.

(45) Fedeles, B. I.; Singh, V.; Delaney, J. C.; Li, D.; Essigmann, J. M. The AlkB Family of Fe(II)/α-Ketoglutarate-Dependent Dioxygenases: Repairing Nucleic Acid Alkylation Damage and Beyond. *J. Biol. Chem.* **2015**, *290* (34), 20734–20742. https://doi.org/10.1074/jbc.R115.656462.

(46) Falnes, P. Ø.; Johansen, R. F.; Seeberg, E. AlkB-Mediated Oxidative Demethylation Reverses DNA Damage in Escherichia Coli. *Nature* **2002**, *419* (6903), 178–182. https://doi.org/10.1038/nature01048.

(47) Trewick, S. C.; Henshaw, T. F.; Hausinger, R. P.; Lindahl, T.; Sedgwick, B. Oxidative Demethylation by Escherichia Coli AlkB Directly Reverts DNA Base Damage. *Nature* **2002**, *419* (6903), 174–178. https://doi.org/10.1038/nature00908.

(48) Aas, P. A.; Otterlei, M.; Falnes, P. O.; Vågbø, C. B.; Skorpen, F.; Akbari, M.; Sundheim, O.; Bjørås, M.; Slupphaug, G.; Seeberg, E.; et al. Human and Bacterial Oxidative Demethylases Repair Alkylation Damage in Both RNA and DNA. *Nature* **2003**, *421* (6925), 859–863. https://doi.org/10.1038/nature01363.

(49) Falnes, P. Ø.; Bjørås, M.; Aas, P. A.; Sundheim, O.; Seeberg, E. Substrate Specificities of Bacterial and Human AlkB Proteins. *Nucleic Acids Res.* **2004**, *32* (11), 3456–3461. https://doi.org/10.1093/nar/gkh655.

(50) Roy, T. W.; Bhagwat, A. S. Kinetic Studies of Escherichia Coli AlkB Using a New Fluorescence-Based Assay for DNA Demethylation. *Nucleic Acids Res.* **2007**, *35* (21), e147. https://doi.org/10.1093/nar/gkm1031.

(51) Zhu, C.; Yi, C. Switching Demethylation Activities between AlkB Family RNA/DNA Demethylases through Exchange of Active-Site Residues. *Angew. Chem. Int. Ed. Engl.* **2014**, *53* (14), 3659–3662. https://doi.org/10.1002/anie.201310050.

CHAPTER 3

HYDROLYSABLE TANNINS ARE IRON CHELATORS THAT INHIBIT DNA REPAIR ENZYME ALKBH2

[SUBMITTED TO CHEMICAL RESEARCH IN TOXICOLOGY]

BY

Fangyi Chen,[†, ±, ‖] Qi Tang,[†, ±] Hang Ma,[†,±] Ke Bian,[†] Navindra P. Seeram,[†] and Deyu Li[*,†]

[†]Department of Biomedical and Pharmaceutical Sciences, College of Pharmacy, University of Rhode Island, Kingston, Rhode Island 02881, United States.

[‖] Xiamen University, 4221 Xiang An South Road, Xiang An District, Xiamen, Fujian 361102, P.R. China (F.C.).

[±]F.C., Q.T. and H.M. contributed equally to this work.

[*]Corresponding Author

E-mail: deyuli@uri.edu

ABSTRACT

Hydrolysable tannins are a class of polyphenolic compounds commonly found in natural products. In this work, we studied the in vitro inhibitory mechanism of six molecules in this class on ALKBH2, an Fe(II)/α-ketoglutarate-dependent DNA repair enzyme in the AlkB family. We determined the IC50 values of these compounds on the repair of 3-methylcytosine, the prototypical substrate of ALKBH2. A structure-activity relationship was also observed between the strength of inhibition and the number of galloyl moieties in a molecule. In addition, we found that the inhibition by this class of polyphenolic compounds on ALKBH2 occur through an iron-chelating mechanism.

INTRODUCTION

Hydrolysable tannins including ellagitannins and gallotannins are a class of polyphenolic compounds that are commonly found in variety of fruits and plants, such as berries and oak barks.[1,2] These molecules share a general structural motif of galloyl group(s) and they have been shown to exhibit many health benefits including antioxidative, antiproliferative, and neuroprotective effects.[3] Recent studies have revealed the involvement of hydrolysable tannins, as iron chelators, in mitigating iron-overload induced hepatotoxicity and neurological disorders.[4] Despite the positive effects, concerns have been raised on the safety of overconsumption of natural products containing hydrolysable tannins, especially with the lack of evidence in the relationship between cellular uptake of these compounds and activity of many iron-related enzymes. There are about 80 Fe(II)/α-ketoglutarate (αKG)-dependent enzymes in human body, including TET, JmjC and AlkB family proteins.[5] In this paper, we

focus on the DNA repair enzyme ALKBH2 in the AlkB family, which use Fe(II)/αKG-dependent mechanism to repair alkyl nucleic acid lesions induced by endogenous and environmental alkylating agents (Figure 1a).[6,7]

The AlkB family proteins include nine human homologs (ALKBH1-8 and FTO), among them ALKBH2 has been identified as the major repair enzyme to lesions in both ss- and ds-DNA, and ALKBH3 has activity mainly on lesions in ss-DNA.[5,6] A number of alkyl substrates have been reported for the AlkB enzymes including 3-methylcytosine (m3C), 1-methlyadenine (m1A), 1,N6-ethenoadenine, 3,N4-ethenocytosine, and other alkyl adducts on the DNA bases. Among these adducts, m3C and m1A have been demonstrated as the best substrates of AlkB.[6,8] Considering the important role of the AlkB enzymes in maintaining genome integrity and the previous observations of hydrolysable tannins as metal chelators,, we questioned whether hydrolysable tannins could affect the activity of the AlkB proteins by chelating the cofactor Fe(II) ion. In this study, we investigated the potential of six hydrolysable tannins (structures 2 to 7, Figure 1b), as a new class of inhibitor to the AlkB enzymes. We extracted and purified a series of ellagitannins and gallotannins to evaluate their inhibitory effects on the repair efficiency of m3C by the ALKBH2 protein by removing the Fe(II) ion cofactor in the enzymatic reaction. In addition, we observed a structure-activity relationship between the strength of inhibition and the number of galloyl groups in these hydrolysable tannins. Our results offer a new perspective into the potential effect of overtaking hydrolysable tannins that could lead to the inhibition of iron-dependent enzymes, which will be confirmed by in vivo studies with the consideration of bioavailability and metabolism.

EXPERIMENTAL PROCEDURES

DNA synthesis. Sixteen-mer oligonucleotides with the sequence of 5'-GAAGACCTXGGCGTCC-3', where X designates 3-methylcytosine, were synthesized by applying automated solid-phase phosphoramidite chemistry on a Mermade-4 DNA synthesizer.[9] The concentration of oligos was measured by NanoDrop (Thermo Scientific) under UV absorbance of 260 nm. The extinction coefficient (ε) of 3-methylcytosine lesion is calculated as its unmodified counterpart due to the negligible value difference in a 16mer context. All the oligos were purified by reverse-phase HPLC with a Phenomenex Luna Semi-Preparative (10 × 250 mm, 5 μm) C18 column. The solvent A was 100 mM 1:1 triethylamine-acetic acid (TEAA) and B was 100% acetonitrile. The oligos were characterized by LC-ESI-TOF-MS (AB Sciex) under negative ionization mode, the liquid chromatographic separation was conducted using a Phenomenex Luna C18 column (4.6 × 100 mm; 5 μm) at a flow rate of 0.4 mL/min. Solvent A was 10 mM ammonium acetate in water, and solvent B was 100% acetonitrile.

Protein purification. The ALKBH2 protein was expressed and purified as previously described.[10] Briefly, the ALKBH2 gene was cloned into the pET28a+ vector and then transformed into E. coli BL21 (DE3) pLysS cells for expression. The his-tagged proteins were purified by affinity column chromatography using FPLC (GE healthcare). Thrombin was added to digest His-tagged proteins overnight followed by further purification using ion-exchange chromatography. The purified ALKBH2 proteins were stored in ALKBH buffer containing 300 mM NaCl, 10% glycerol, 50

mM N-[tris(hydroxymethyl)methyl]-3-aminopropanesulfonic acid and 1 mM 2-mercaptoethanol under pH 8.0.

Enzymatic assay and Inhibition mechanism study. For the enzyme inhibition studies, 5 μM of single-stranded m3C DNA oligonucleotides were incubated with 1.25 μM of ALKBH2 proteins in 46.5 mM HEPES buffer (pH8.0) containing 50 μM αKG, 5 μM Fe (NH4)2(SO4)2 and 1.86 mM ascorbic acid in a 20 μL reaction volume at 37°C for 1 hour.2 To test the IC50 values, the concentrations of hydrolysable tannins used were optimized individually to achieve the inhibition percentages between range of 0% to 100%. For the inhibition mechanism studies, the concentration of αKG was fixed to 93 μM, and different concentrations of Fe (NH4)2(SO4)2 (5, 10, 20, 40, 80 μM) were used for Fe(II) chelator test. In contrast, varying concentrations of αKG (50, 100, 200, 400, 800 μM) and fixed 10 μM of Fe (NH4)2(SO4)2 were used for the α-ketoglutarate competitor test. The enzyme activity was inactivated by adding 10mM EDTA followed by heating up to 95°C for 5 min. The analyses were performed by HPLC and all data represent the mean (± standard deviation, SD) of three independent experiments.

HPLC analysis. m3C and its repaired product C were analyzed by using a DNApac PA-100 anion-exchange column (4 x 250 mm, 13 μm) with solvent A as water and solvent B as 1.5 M ammonium acetate in water. The flow rate was at 1.0 mL/min. A solvent gradient was carried out under the following conditions: 60% of B for 6 min, 60%-80% for 0.5 min, 80% for 1 min, 80%-60% for 0.5 min, 60% for 4 min.

Materials. Rhein (1), gallic acid (2), ellagic acid (3) and tannic acid (7) were purchased from Sigma-Aldrich Chemical Co. (St. Louis, MO, USA). Ellagitannin (punicalagin 6) was isolated from a commercially available pomegranate extract (PomellaTM) provided by Verdure Sciences (Noblesville, IN, USA) using our previously reported methods.[11] Gallotannins including ginnalin A (4) and pentagalloyl glucose (5) were isolated from a red maple leaves extract and sumac fruit extract, respectively, as we previously reported.[12,13]

RESULTS AND DISCUSSION

To test the inhibitory effects of hydrolysable tannins on ALKBH2, 16mer oligonucleotides bearing a site-specific m3C adduct, the primary substrate of ALKBH2, were synthesized by using solid-phase phosphoramidite chemistry.[9,14] The recombinant human ALKBH2 protein was expressed and purified according to previous studies.[10,15] Then the repair efficiency of ALKBH2 on m3C was assessed. For the enzymatic reactions, the m3C-containing oligonucleotides were incubated with ALKBH2 in the presence of two cofactors (αKG and Fe(II) ion) and other necessary reagents and buffers (see Experimental section in SI for details), and the repair ratio of m3C was quantified by HPLC after reaction. We used 5.0 μM DNA and 1.25 μM ALKBH2 for a typical enzymatic reaction to test the inhibition of different compounds, these concentrations are similar to the conditions previously reported.[15]

The IC50 values of the six hydrolysable tannins on inhibition of ALKBH2 were determined. We prepared these compound as previously reported.[11,12,16] Rhein (compound 1, Figure 1b), a well-known inhibitor to the AlkB family enzymes was served as a positive control.[17] The concentrations of rhein varied from 0 to 150 μM

and the IC50 was found as 42.0 μM (Figure S1), which is comparable to the values on inhibiting AlkB and FTO reported by Li et al.[17] For the hydrolysable tannins, the concentrations were tested between 0 and 80 μM (Figures S2 to S7). The IC50 values of compound 2 to 7 were found from 3.9 to 38.2 μM (Table1 and Figures S2 to S7). Interestingly, based on the structures of these hydrolysable tannins, we found a correlation of their structural features to their IC50 values: as the number of galloyl group in the molecule increases, the IC50 value decreases (Table 1). For example, the one galloyl unit containing compound 2 has IC50 38.2 μM, the five galloyl unit containing compound 5 has IC50 18.9 μM, and the ten galloyl unit containing compound 7 has IC50 3.9 μM. Compounds 3 and 4 have two galloyl units and they have similar IC50 values (25.5 and 22.6 μM, respectively). These results indicate that the basic galloyl unit of these derivatives are potentially the structural motif responsible for the inhibitory effects.

We further explored the inhibition mechanism of compounds 1 to 7 on the DNA repair by ALKBH2. Previously, rhein has been reported as an inhibitor to AlkB and FTO by competitively replacing αKG in the active site.[17,18] And hydrolysable tannins, including tannic acid (compound 7), have been reported to ameliorate Fe(II) ion and reactive oxygen species induced neurological diseases.[19] In addition, tannic acid has been reported to mitigate iron-overload by its chelating ability through the neighboring phenol hydroxyl groups.[4] We wanted to investigate the inhibitory mechanisms of the hydrolysable tannins and chose tannic acid because of its reported activities and the most galloyl groups possessed in the molecule.

We varied the concentration of αKG and Fe(II) ion in the reactions without any inhibitor, and with rhein or compound 7. For the reactions without the inhibitor, small variations on the reactivity were found with αKG changing from 50 to 800 μM (89% to 99%, control in Figure 2a) and Fe(II) ion changing from 5 to 80 μM (staying around 93%, control in Figure 2b). First we tested the reactivity change with different concentrations of αKG with a fixed Fe(II) ion concentration of 10 μM. When 50 μM of rhein was added to the reaction media, we found the repair ratio decreased to 30% with 50 μM αKG. As the amount of αKG increased up to 800 μM, the repair efficiency was recovered in a concentration-dependent manner (Figure 2a). Then we tested the reactivity change with different concentrations of Fe(II) ion (from 5 to 80 μM) with a fixed αKG concentration of 93 μM. The reactions with 50 μM rhein all had repair efficiency around 36% and showed no significant change even with the increment of Fe(II) ion to 80 μM (Figure 2b). These observations are consistent with previously reported data of rhein and show that it is a competitive inhibitor to αKG but not an Fe(II) ion chelator.[17,18] For compound 7 under variable concentrations of αKG, all of the reactions with 2 μM of it had a decreased repair efficiency around 25% and did not vary significantly with the increment of αKG even to 800 μM (Figure 2a, Fe(II) ion kept at 10 μM). For the reactions with varying concentrations of Fe(II) ion, the reactivity was decreased to 5% with 5 μM Fe(II) ion and was restored in a concentration-dependent manner as Fe(II) ion stepwisely increased up to 80 μM (Figure 2b, αKG kept at 93 μM). These data indicate that compound 7 is an Fe(II) ion chelator but not a competitive inhibitor to αKG.

In this study, we investigated hydrolysable tannins as a new class of natural product inhibitors for ALKBH2, the major human DNA repair enzyme in the AlkB family. Hydrolysable tannins have been identified as antioxidative, antiproliferative and neuroprotective agents. Here, we discovered that these compounds are Fe(II) ion chelators that can decrease the level of free Fe(II) ion, which is critical for the catalytic activity of the AlkB family enzymes. At a concentration higher than the Fe(II) ion level, the reactivity of the AlkB enzymes might be inhibited. It is possible that the overconsumption of hydrolysable tannins may affect iron homeostasis in the body leading to the dysfunction of iron-dependent proteins, such as the AlkB homologs and other Fe(II)/αKG-dependent enzymes. Further studies are necessary to confirm these observations.

FIGURES

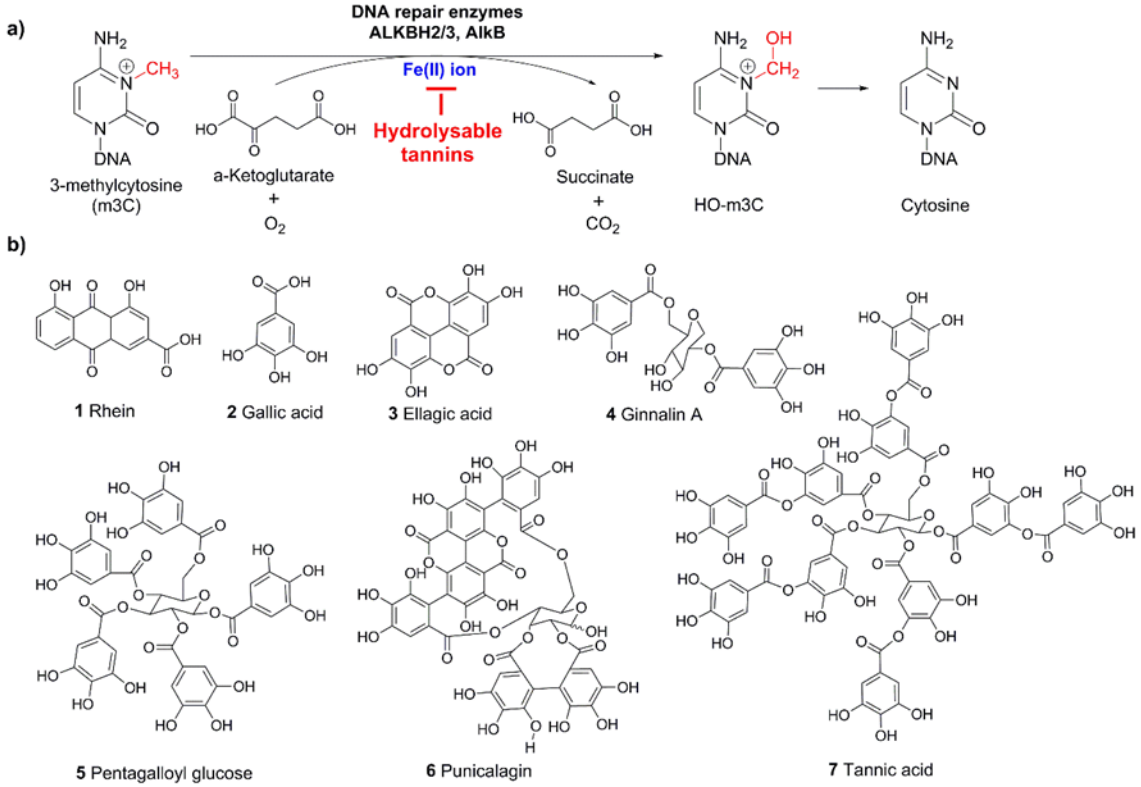

Figure 1. Repair mechanism of the AlkB family enzymes and structures of the hydrolysable tannins and rhein.

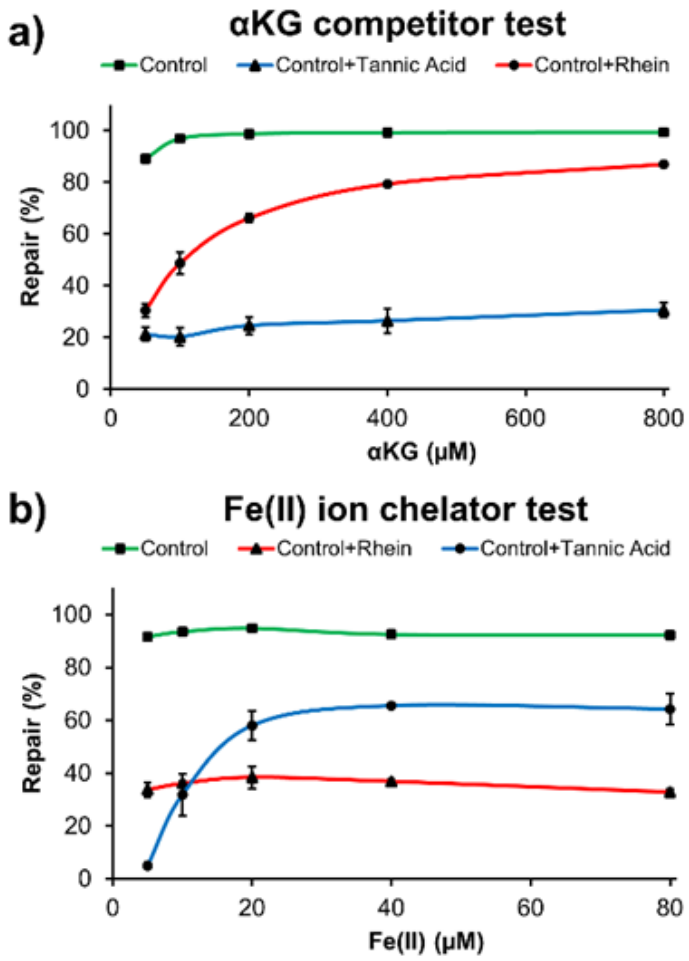

Figure 2. Inhibition of ALKBH2 repair reactions by tannic acid and rhein (control: no inhibitor added).

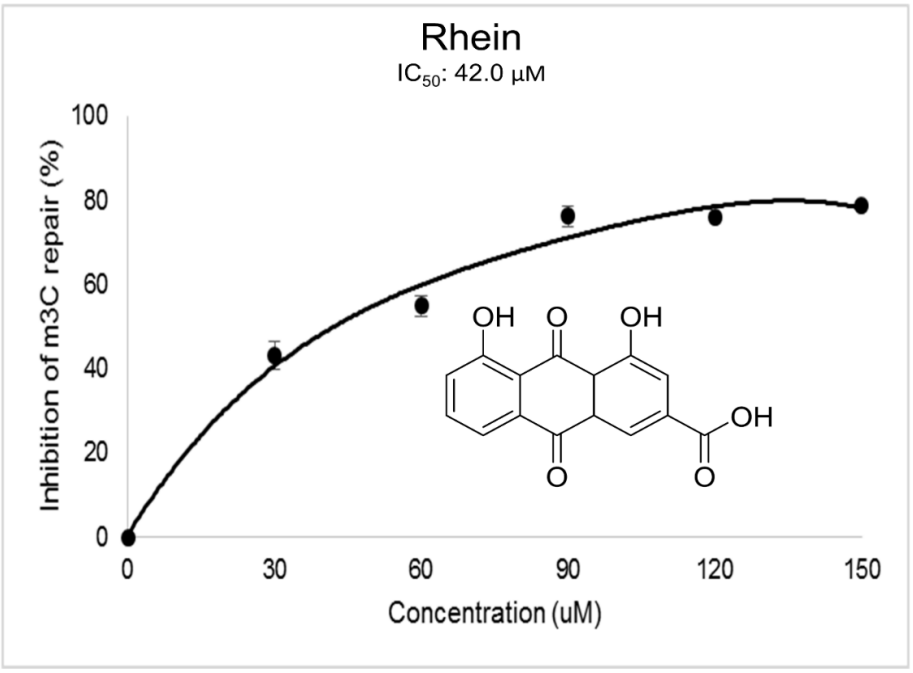

Figure S1. Data for determining the IC_{50} value of compound **1** (Rhein).

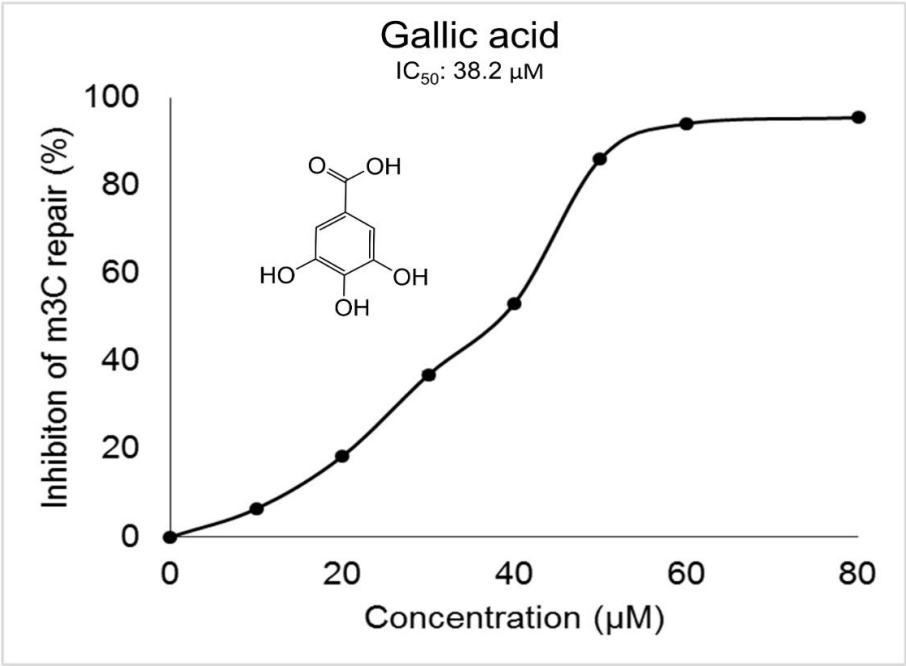

Figure S2. Data for determining the IC_{50} value of compound **2** (gallic acid).

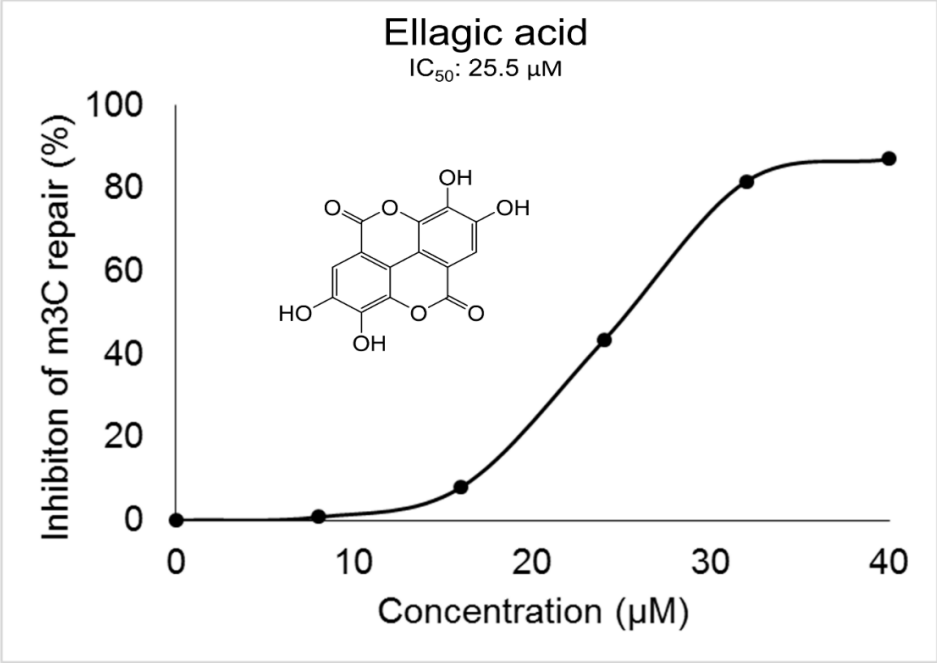

Figure S3. Data for determining the IC_{50} value of compound **3** (ellagic acid).

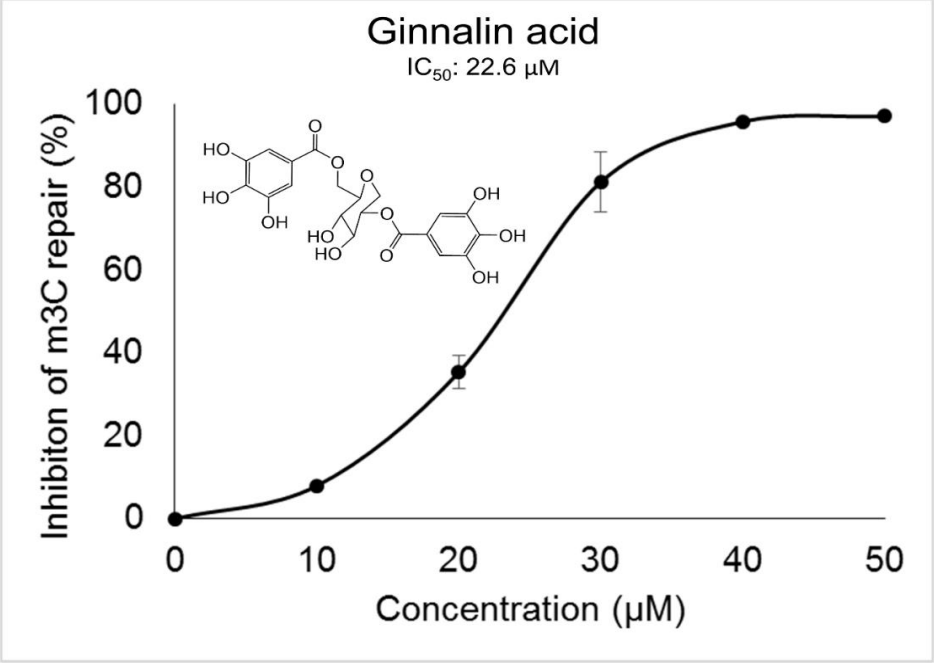

Figure S4. Data for determining the IC_{50} value of compound **4** (ginnalin acid).

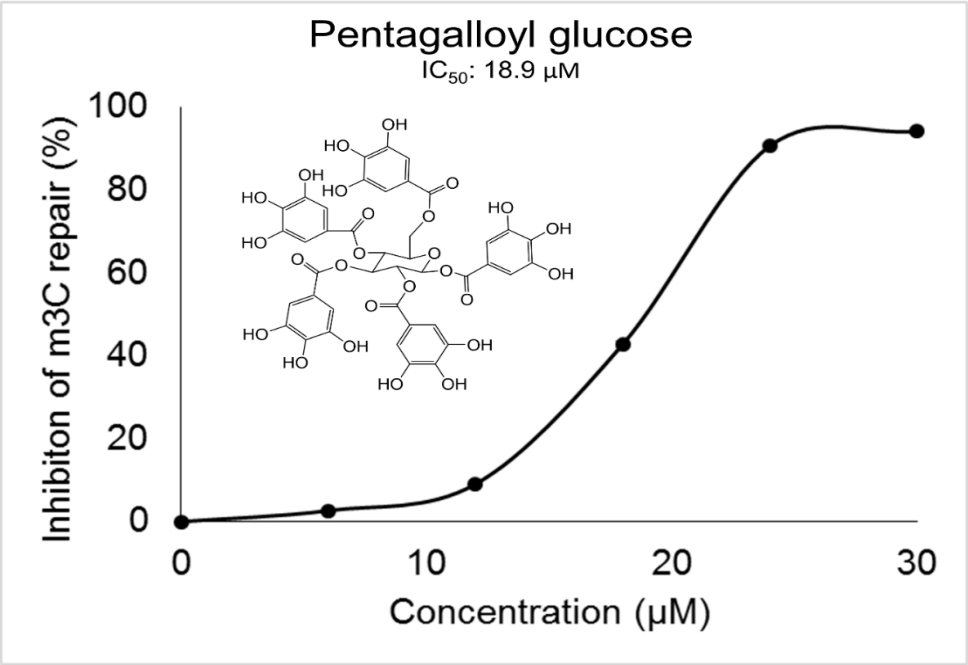

Figure S5. Data for determining the IC_{50} value of compound **5** (pentagalloyl glucose).

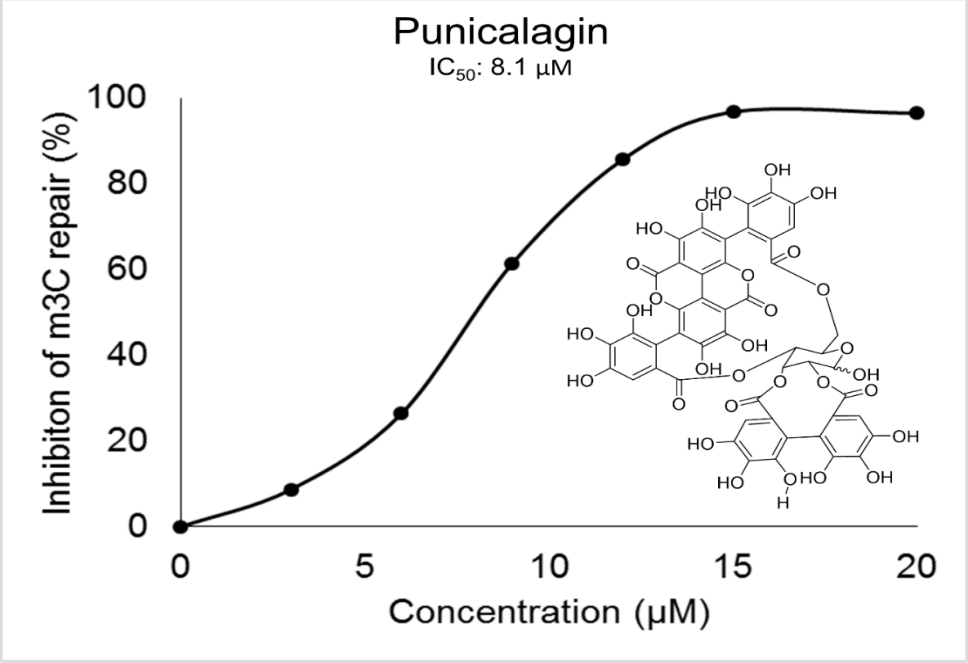

Figure S6. Data for determining the IC_{50} value of compound **6** (punicalagin).

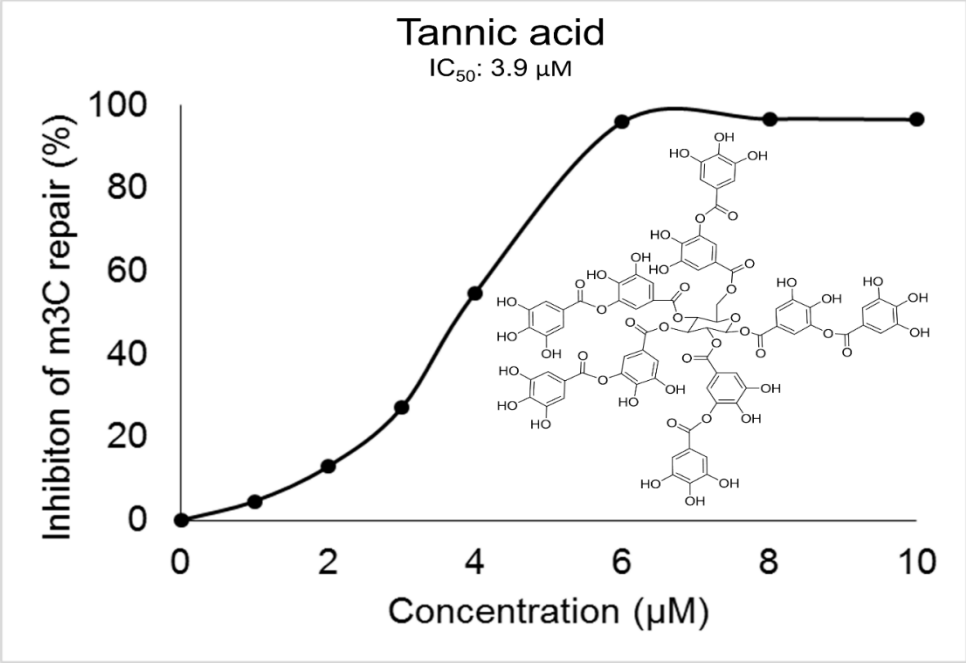

Figure S7. Data for determining the IC_{50} value of compound **7** (tannic acid).

TABLES

Table 1. IC50 values of hydrolysable tannins inhibiting the ALKBH2 enzyme.

compound	IC_{50} (µM)	galloyl units
2	38.2	1
3	25.5	2
4	22.6	2
5	18.9	5
6	8.1	6
7	3.9	10

ACKNOELEDGEMENTS

The authors want to thank the RI-INBRE program, its director Prof. Bongsup Cho, and staff Dr. Al Bach, Kim Andrews, and Patricia Murray for their kind help. This work was supported by an Institutional Development Award from the National Institute of General Medical Sciences of the National Institutes of Health under grant number P20 GM103430. This work was also supported by National Institutes of Health under grant numbers R15 CA213042 and R01 ES028865 (to D.L.).

REFERENCES

(1) Okuda, T.; Ito, H. Tannins of Constant Structure in Medicinal and Food Plants—Hydrolyzable Tannins and Polyphenols Related to Tannins. *Molecules* **2011**, *16* (3), 2191–2217. https://doi.org/10.3390/molecules16032191.

(2) Arapitsas, P. Hydrolyzable Tannin Analysis in Food. *Food Chem* **2012**, *135* (3), 1708–1717. https://doi.org/10.1016/j.foodchem.2012.05.096.

(3) Chung, K. T.; Wong, T. Y.; Wei, C. I.; Huang, Y. W.; Lin, Y. Tannins and Human Health: A Review. *Crit Rev Food Sci Nutr* **1998**, *38* (6), 421–464. https://doi.org/10.1080/10408699891274273.

(4) Basu, T.; Panja, S.; Shendge, A. K.; Das, A.; Mandal, N. A Natural Antioxidant, Tannic Acid Mitigates Iron-Overload Induced Hepatotoxicity in Swiss Albino Mice through ROS Regulation. *Environ. Toxicol.* **2018**, *33* (5), 603–618. https://doi.org/10.1002/tox.22549.

(5) *2-Oxoglutarate-Dependent Oxygenases*; Hausinger, R. P., Schofield, C. J., Eds.; RSC metallobiology; Royal Society of Chemistry: Cambridge, UK, 2015.

(6) Fedeles, B. I.; Singh, V.; Delaney, J. C.; Li, D.; Essigmann, J. M. The AlkB Family of Fe(II)/α-Ketoglutarate-Dependent Dioxygenases: Repairing Nucleic Acid Alkylation Damage and Beyond. *J. Biol. Chem.* **2015**, *290* (34), 20734–20742. https://doi.org/10.1074/jbc.R115.656462.

(7) Yi, C.; He, C. DNA Repair by Reversal of DNA Damage. *Cold Spring Harb. Perspect. Biol.* **2013**, *5* (1), a012575. https://doi.org/10.1101/cshperspect.a012575.

(8) Sedgwick, B. Repairing DNA-Methylation Damage. *Nat. Rev. Mol. Cell Biol.* **2004**, *5* (2), 148–157. https://doi.org/10.1038/nrm1312.

(9) Chen, F.; Tang, Q.; Bian, K.; Humulock, Z. T.; Yang, X.; Jost, M.; Drennan, C. L.; Essigmann, J. M.; Li, D. Adaptive Response Enzyme AlkB Preferentially Repairs 1-Methylguanine and 3-Methylthymine Adducts in Double-Stranded DNA. *Chem. Res. Toxicol.* **2016**, *29* (4), 687–693. https://doi.org/10.1021/acs.chemrestox.5b00522.

(10) Chen, F.; Bian, K.; Tang, Q.; Fedeles, B. I.; Singh, V.; Humulock, Z. T.; Essigmann, J. M.; Li, D. Oncometabolites D- and l-2-Hydroxyglutarate Inhibit the AlkB Family DNA Repair Enzymes under Physiological Conditions. *Chem. Res. Toxicol.* **2017**, *30* (4), 1102–1110. https://doi.org/10.1021/acs.chemrestox.7b00009.

(11) Yuan, T.; Ma, H.; Liu, W.; Niesen, D. B.; Shah, N.; Crews, R.; Rose, K. N.; Vattem, D. A.; Seeram, N. P. Pomegranate's Neuroprotective Effects against Alzheimer's Disease Are Mediated by Urolithins, Its Ellagitannin-Gut Microbial Derived Metabolites. *ACS Chem Neurosci* **2016**, *7* (1), 26–33. https://doi.org/10.1021/acschemneuro.5b00260.

(12) Ma, H.; Liu, W.; Frost, L.; Kirschenbaum, L. J.; Dain, J. A.; Seeram, N. P. Glucitol-Core Containing Gallotannins Inhibit the Formation of Advanced Glycation End-Products Mediated by Their Antioxidant Potential. *Food Funct* **2016**, *7* (5), 2213–2222. https://doi.org/10.1039/c6fo00169f.

(13) Ma, H.; Liu, W.; Frost, L.; Wang, L.; Kong, L.; Dain, J. A.; Seeram, N. P. The Hydrolyzable Gallotannin, Penta-O-Galloyl-β-D-Glucopyranoside, Inhibits the Formation of Advanced Glycation Endproducts by Protecting Protein Structure. *Mol Biosyst* **2015**, *11* (5), 1338–1347. https://doi.org/10.1039/c4mb00722k.

(14) Tang, Q.; Cai, A.; Bian, K.; Chen, F.; Delaney, J. C.; Adusumalli, S.; Bach, A. C.; Akhlaghi, F.; Cho, B. P.; Li, D. Characterization of Byproducts from Chemical Syntheses of Oligonucleotides Containing 1-Methyladenine and 3-Methylcytosine. *ACS Omega* **2017**, *2* (11), 8205–8212. https://doi.org/10.1021/acsomega.7b01482.

(15) Bian, K.; Chen, F.; Humulock, Z. T.; Tang, Q.; Li, D. Copper Inhibits the AlkB Family DNA Repair Enzymes under Wilson's Disease Condition. *Chem. Res. Toxicol.* **2017**, *30* (10), 1794–1796. https://doi.org/10.1021/acs.chemrestox.7b00230.

(16) Ma, H.; Liu, W.; Frost, L.; Wang, L.; Kong, L.; Dain, J. A.; Seeram, N. P. The Hydrolyzable Gallotannin, Penta-O-Galloyl-β-D-Glucopyranoside, Inhibits the Formation of Advanced Glycation Endproducts by Protecting Protein Structure. *Mol Biosyst* **2015**, *11* (5), 1338–1347. https://doi.org/10.1039/c4mb00722k.

(17) Li, Q.; Huang, Y.; Liu, X.; Gan, J.; Chen, H.; Yang, C.-G. Rhein Inhibits AlkB Repair Enzymes and Sensitizes Cells to Methylated DNA Damage. *J. Biol. Chem.* **2016**, *291* (21), 11083–11093. https://doi.org/10.1074/jbc.M115.711895.

(18) Chen, B.; Ye, F.; Yu, L.; Jia, G.; Huang, X.; Zhang, X.; Peng, S.; Chen, K.; Wang, M.; Gong, S.; et al. Development of Cell-Active N6-Methyladenosine RNA Demethylase FTO Inhibitor. *J. Am. Chem. Soc.* **2012**, *134* (43), 17963–17971. https://doi.org/10.1021/ja3064149.

(19) Chan, S.; Kantham, S.; Rao, V. M.; Palanivelu, M. K.; Pham, H. L.; Shaw, P. N.; McGeary, R. P.; Ross, B. P. Metal Chelation, Radical Scavenging and Inhibition of $A\beta_{42}$ Fibrillation by Food Constituents in Relation to Alzheimer's Disease. *Food Chem* **2016**, *199*, 185–194. https://doi.org/10.1016/j.foodchem.2015.11.118.

CHAPTER 4

GLYCATION BY GLUCOSE 6-PHOSPHATE INDUCES G TO T MUTATION AND SINGLE NUCLEOTIDE DELETION IN CELL

[MANUSCRIPT IN PREPARATION]

BY

Qi Tang,[†] Ke Bian,[†] Fangyi Chen,[†] and Deyu Li[*,†]

[†]Department of Biomedical and Pharmaceutical Sciences, College of Pharmacy, University of Rhode Island, Kingston, Rhode Island 02881, United States.

[*]Corresponding Author

E-mail: deyuli@uri.edu

ABSTRACT

Reducing sugars non-enzymatically glycate proteins and nucleic acids to form glycation adducts. Previously, glucose and glucose-6-phosphate have been shown to glycate lysine residues of protein to form Amadori adducts. As a glycating agent with multiple-fold higher reactivity and much more abundant cellular concentration than glucose, glucose-6-phosphate has been proposed to glycate DNA and cause genotoxicity. Due to the unstable nature of the Amadori adducts and the presence of cellular DNA repair pathways, it is challenging to study their toxic and mutagenic properties in cell. Therefore, the specific biological outcomes of these adducts remain unclear. In this work, we used chemical and genetic approaches to study the biological consequences, such as replication block and mutations, of Amadori DNA adducts in a site-specific manner. Our data showed that Amadori DNA adducts arising from glucose-6-phosphate on dG can induce G to T and single nucleotide deletion, which may lead to devastating biological consequences if left unrepaired.

INTRODUCTION

Reducing sugars and their metabolic derivatives (e.g. glucose, glucose 6-phosphate, methylglyoxal, and glyoxal) can nonenzymatically react with free amino groups of biomacromolecules, such as proteins and nucleic acids, to induce glycation.[1–3] Among these cellular glycating agents, methylglyoxal and glyoxal are byproducts of sugar metabolism which are considered as toxins, whereas glucose or glucose 6-phosphate (Glu-6-P) are thought to be non-toxic because they are essential metabolic intermediates to supply energy and precursors for biosynthetic pathways in nearly all organisms. Although D-glucose has a high concentration in blood, it is

mostly phosphorylated to Glu-6-P after it entering cell (Fig. 1). Thus, Glu-6-P is the most abundant reducing sugar in cell (~ mM level) which can be further elevated under abnormal metabolic conditions.[1,4]

Previously, Glu-6-P and glucose have been shown to glycate lysine residues of protein to form fructosamine 6-phosphate (FN6P) and fructosamine products, respectively.[5,6] And these adducts are involved in the pathogenesis of many diseases, such as aging and cardiovascular diseases due to accumulation of the unrepaired glycated-proteins.[2,2] Recent studies showed that protein FN6P and fructosamine adducts can be repaired by MDP-1/FN3K mechanism, which has been hypothesized as the major detoxification pathway for Glu-6-P induced protein glycation.[7,8] As a glycating agent with multiple-fold higher activity than glucose, Glu-6-P has been demonstrated to glycate DNA and lead to cytotoxicity and mutagenesis. DNA glycation has been shown to lead to increased cytotoxicity, strand breaks, and mutation frequency that are thought to be associated with cancer, neurological disorders, and aging related diseases.[9,2] The extent of DNA alternation appears to be dependent on the type of nucleobase involved in the modification. Among the three amino group-containing nucleobases, guanine is most vulnerable to glycating agents; and N^2-dG adducts are the primary source of DNA glycation products.[2,5]

Glu-6-P glycates DNA with the formation of unstable Schiff base, which rapidly undergoes structural rearrangement and thereby generate an acyclic Amadori product (N^2-Fructosamine 6-Phosphate-Deoxygunosine, N^2-FN6P-dG), the acyclic N^2-FN6P-dG adduct undergoes further ring closure at O-5 with the C-2 carbonyl group to generate more stable α and β configurations (Fig. 1). Previous cellular study reported

that elevated intracellular Glu-6-P levels could increase plasmid mutations in glycolytic mutant bacteria.[10] Also, Bucala et al. examined the mutagenic effect of Glu-6-P-induced adducts in pBR32 DNA plasmid in *E. coli*, and a number of mutant plasmids were observed to have undergone gross DNA alterations, including mutations, insertions and deletions, as well as the development of multiple species originating from a single cell.[11] These studies have demonstrated that DNA glycation by Glu-6-P can cause damages to genome and lead to mutagenicity. However, with the complexity of the formation and unstable nature of Glu-6-P glycation adducts as well as the potential cellular repair mechanisms involved, it is challenging to study the mutagenic properties of these adducts. However, with the complexity of the formation and unstable nature of Glu-6-P glycation adducts as well as the potential cellular repair mechanisms involved, it is challenging to study the mutagenic properties of these adducts. For these reasons, the exact biological consequences of glycation DNA adducts from Glu-6-P are still unclear.

In this paper, we used chemical and genetic methods to study the biological consequences, such as replication block and mutagenesis, of these glycation adducts in a site-specific manner. We chemically synthesized N^2-FN6P-dG adducts and site-specifically incorporated them into single-stranded M13 vectors, ant the cytotoxic and mutagenic properties of the adducts were determined in *E. coli* cells. Our results, for the first time, revealed that N^2-FN6P-dG adducts cause strong DNA replication block as well as significant amount of G → T and single nucleotide deletion mutations in cell. These observations indicate that FN6P DNA glycation products arising from Glu-

6-P could lead to genome instability if left unrepaired, thereby affecting the genome integrity of organisms.

EXPERIMENTAL PROCEDURES

Chemical Synthesis and Characterization of N^2-FN6P-dG Glycation Products. Thirteen-mer unmodified oligonucleotides with the sequence 5'-TTTTTTGTTTTTT-3' were synthesized by using solid-phase phosphoramidite chemistry on a DNA synthesizer (Bioautomation). The sequence was designed to harbor only one guanine base to ensure the site-specific incorporation of N^2-FN6P-dG glycation adduct. A concentration of 200 μM 13mer oligos was incubated with 1 M of D-Glucose 6-phosphate (Roche) under physiological conditions (pH 7.4 in PBS buffer at 37 °C) for 14 days, two major N^2-FN6P-dG products were observed. Studies have shown that the glycation process was accelerated under elevated pH and temperature [1,12]. An optimized reaction condition (pH 9.5 in 0.1 M sodium borate buffer at 80 °C) was used to shorten the reaction time (12 h) and increase the yield of glycation products, two N^2-FN6P-dG products were observed with HPLC retention time and MW matching the two glycation products formed under physiological conditions. The resulting two N^2-FN6P-dG glycation products were separated and purified by a DNAPac PA100 anion-exchange oligonucleotide column (13 μm, 22 x 250 mm, Thermo Scientific) with two solvents. Solvent A was water, and solvent B was 1.5 M ammonium acetate. The purified glycation products were characterized by high-resolution triple-TOF MS (AB Sciex).

Construction of M13 Vectors Containing Site-Specific Incorporation of N^2-FN6P-dG Adducts. The M13mp7(L2) viral genomes were prepared with site-

specific incorporation of N^2-FN6P-dG or adduct-free N^2-dG by following the procedures adopted from Delaney and Essigmann.[13] Specifically, 1000 pmol of 13mer adduct carrying ss-oligos (5'-TTTTTTXTTTTTT-3', where "X" designates N^2-FN6P-dG or N^2-dG) and 1050 pmol of 21mer barcode ss-oligos (1st barcode sequence: 5'-TACCGTCGBBBCGCGCATGCA-3', and 2nd barcode sequence: 5'-TCTCGAGTGBBBCGTCAGCAC-3', where BBB serves as the barcode location for next-generation sequencing purpose that was not applied in this work) were 5'-phosphorylated in a 50 μl solution containing 1mM ATP, 1×T4 polynucleotide kinase buffer, and 30 units of polynucleotide kinase at 37 °C for 1 h. The 5'-phosphorylated 13mer strand and two 21mer strands were annealed with two scaffolds (1100pmol of 15mer Scaffold I: 5'-GCGCGTACGTAAAAA-3', and 1100pmol of 14mer Scaffold II: 5'-AAAAAAGAGCTCAC-3') and further elongated into a 55mer strand by using 1200 units of T4 DNA ligase at 16 °C for 3 h. The ligated 55mer oligos were purified by 15% PAGE-urea gel and the concentration was quantified by UV absorbance measurement at 260 nm. The ligated 55mer oligos (75 pmol) were annealed with 30 pmol of ss-M13mp7(L2) vectors (linearized by XX units of EcoRI at XX °C for XX h) in the presence of two scaffolds (48 pmol of 28mer Scaffold III: 5'-CGATACTGGTACTAAGTCACATGGCAGC-3', and 48 pmol of 29mer Scaffold IV: 5'-GCAGTCGTGTTAAGTGACCGGCAGCAAAA-3') and then ligated by T4 ligase at 16 °C for 12 h. 20units λ exonuclease and 30 units of RecJf exonuclease were subsequently added to the resulting mixture and incubated at 37 °C for 5 h. The mixture was extracted with phenol/chloroform/isoamyl alcohol (25/24/1, vol), and the aqueous phase was precipitated by 100% ethanol and the DNA pellet (genome) was

resuspended in nuclease-free water. The constructed genomes were normalized against an adduct-free competitor genome ligated with a 58mer competitor strand (by inserting a 16mer ss-oligo: 5'-TTTTTTGTTTTTTTTT-3').

Transfection of *E. coli* Cells with M13 Vectors Containing N^2-FN6P-dG adducts. The electrocompetent AB1157 *E. coli* cells were electroporated (at 2.5kV) with M13 genomes containing N^2-FN6P-dG or adduct-free N^2-dG which were pre-mixed with the competitor genomes at a ratio of 4:1. The *E. coli* cells were then transferred into 1 ml SOC medium and incubated at 37 °C for 1 h followed by transferring into 10 ml LB medium for an additional incubation of 6 h, after which the resulting phages were collected from the supernatant of culture medium by centrifuging at 4700 rpm for 15 min. The phages were further amplified by SCS110 *E. coli* cells to increase the progeny genome percentage over adduct-carrying genomes. The final amplified phages were recovered by centrifuging at 4700 rpm for 15 min, and the M13 genomes were extracted with a QIAprep 2.0 spin column by following manufacturer's guidance.

LC-MS Based Bypass Efficiency and Mutation Frequency Assays. A competitor gapped plasmid (with a ligation of 16mer oligonucleotide 5'-TTTTTTGTTTTTTTTT-3') with three bases longer than the adduct-containing plasmid was used as an internal control. The resulting progeny genomes after replication were extracted according to the instruction of QIAGEN DNA extraction kit. The region of interest in the genome was amplified with two primers (Primer F: 5'-GCTATGACCATGATTCAGTGTACCGTCG-3', Primer R: 5'-AAAACGACGGCC AGTGAATTGTGCTGACG-3') by PCR. The PCR products were purified by

QIAquick PCR Purification kit. For the LC-MS analysis, the PCR fragments were digested with 40 units of Xhol and 30 units of SphI enzymes in 50μL of CutSmart buffer at 37 °C for 3 h, flowed by heating up to 65 °C for 20 min to deactivate the restriction enzymes. The resulting oligonucleotides were desalted and analyzed by LC-MS. The following LC-MS conditions were used: the LC separation was carried out by using a Thermo Acclaim PolarAdvantage II C18 column (3μm, 2.1 x 250mm); the flow rate was 100μL; a 5min of gradient of 25% was used to introduce sample followed by a 30 min of 25%-50% methanol in 400mM HFIP buffer (pH 7.0) for separation. ESI was conducted by using a needle voltage of 4.0 kV in a negative mode. The heater was set at 300 °C. The nebulizer gas was 40 psi; the heater gas was 40 psi; the curtain gas was 25 psi; the declustering potential was −220 V; and the collision energy was −10 V. For the calculation of bypass efficiency, the formula %bypass = (adduct oligo signal/competitor signal)/(non-adduct signal/its competitor signal) was used.

Statistical Analyses. Statistical analysis was conducted by software IBM SPSS Statistics 23 and Microsoft Excel 2016. Statistical significance for data between two groups was characterized with a two-tailed Student's t test. P-values < 0.05 were considered statistically significant.

RESULTS

Chemical Synthesis and Characterization of N^2-FN6P-dG Adducts. To synthesize N^2-FN6P-dG glycation adducts, we first tested the glycation of DNA under physiological conditions (37 °C, and pH 7.4 in PBS buffer) by incubating 200 μM of 13mer oligonucleotide (5'-TTTTTTGTTTTTT-3') with 1 M Glu-6-P for 14 days.

AEX-HPLC analysis have revealed the formation of two major products (peak1 and 2 in Fig. 2b), which were later characterized as α and β anomers of the N^2-FN6P-dG adducts in the 13mer context. However, the yield of the products was low (less than 10%), which did not meet the amount of material required for the subsequent biological studies. As the glycation rate of nucleic acids by reducing sugars was reported to be dependent on two major factors: pH and temperature,[12,14] we optimized the reaction condition by applying pH 9.5 under 80 °C. We monitored the formation of glycation products after 8 hours of incubation, and the results showed that the reaction was greatly accelerated and could be prepared on a large scale (Fig. 2a). We then conducted the purification for the two glycation products by AEX-HPLC (Fig. 3a), and the HPLC chromatogram indicated that the two peaks were well separated (Fig. 3c and 3d). As the glycation process is reversible, we also tested the stability of the two products by incubating them under physiological conditions for 4 days, some of the two products (Fig. 3e for peak 1 and 3f for peak 2) were converted back to the starting material 13mer G (Fig. 3b); these observations are consistent with the previous studies that the DNA glycation of Amadori products are reversible to unmodified DNA bases.[15] In addition, we compared the reactivity of different nucleobases to Glu-6-P under the optimized reaction condition. We quantified the glycation products formation for the 13mer oligos containing dA, dC, dG and dT nucleobase respectively. The results showed that 13mer G exhibited highest reactivity with 23.1% glycation conversion after 8 h, whereas 13mer A and 13mer C had only 1.6% and 0.8%, there was no glycation on dT (Fig. 4).

After synthesis, we observed the formation of two major glycation adducts (Fig 2, and Fig 3a) from the incubation of Glu-6-P with DNA oligos. To identify the MW of the two adducts, the modified oligos were analyzed by high-resolution MS. The 13mer glycated oligo (peak 1 in Fig. 3c) showed a good signal at m/z of 1384.8857 at the -3 charged envelope of mass spectrum (Fig. 5a). The observed m/z matches the theoretical calculation of m/z of 1384.8816 for N^2-FN6P-dG adduct formation at X location of the starting material sequence (5'-TTTTTTXTTTTTT-3'). Similarly, we observed identical MW for the second glycation adduct (peak 2 in Fig. 3d, MS data not shown). These results indicated the two glycation products were α and β anomers from ring closure of Amadori product at O-5 with the C-2 carbonyl group (Fig. 1). To further confirm the identity of the adducts, we conducted nuclease digestion and dephosphorylation of the purified N^2-FN6P-dG adduct-containing oligos into single nucleosides, then analyzed the highly hydrophilic adduct-carrying dG by HILIC-LC-MS. By using a combination of DNase I, phosphodiesterase I, and alkaline phosphatase to digest the 13mer sequence into single nucleosides,[16–18] we were able to observe the digested adduct-modified nucleoside at m/z 430.1647 (positive mode, +H, 430.1574 for theoretical m/z); this species was generated from the enzymatic dephosphorylation of N^2-FN6P-dG into N^2-Fructosamine-dG (Fig. 5b, representative MS data for peak 1 digestion in Fig. 3c). In addition, no thymidine modifications have been observed in the 13mer sequence. These observations together supported that the glycation products were N^2-FN6P-dG adducts from Glu-6-P (Fig. 1).

Replication Block and Mutagenicity Assays of N^2-FN6P-dG Adducts. We next determined how the presence of N^2-FN6P-dG adducts affect DNA replication,

and what biological outcomes are involved in polymerase bypassing these adducts in cell. To this end, we carried out site-specific detection of replication block and mutagenesis adopted from Essigmann lab [13,19] for the two N^2-FN6P-dG adducts. We ligated the two adduct-containing 13mer oligos into single-stranded M13 vector respectively (Fig. 6), and tested the bypass efficiency and mutation frequency of them in *E. coli* cells. The data obtained from quantitative LC-MS analysis showed adduct-1 with a bypass efficiency of only 22% comparing to unmodified guanine as 100%, and adduct-2 with a bypass efficiency of 24% (Fig. 7, top). These results indicated that the N^2-FN6P-dG adducts are strong blocks to DNA replication. We subsequently quantified the mutation frequency of the two adducts, adduct 1 (Fig. 7, bottom) exhibited strong G to T mutation (26%) and -1 deletion of dG base (30%). Similarly, adduct-2 had G to T mutation of 16% and -1 deletion of dG of 36%. These results demonstrated that N^2-FN6P-dG adducts can cause strong mutagenicity.

DISCUSSION

Glu-6-P is an important glycolysis intermediate which constitutes a dominant endogenous reducing sugar metabolite. Previous studies on its cellular glycation outcomes demonstrated that the dysregulation of Glu-6-P is associated with genotoxicity including insertion, deletion, and point mutations[10,11]. However, neither the exact mutation patterns of the glycation nor the specific identity of these glycation adducts were assessed. This is challenging presumably that, first, cellular glycation process is complex due to the many types of glycation adducts formed; second, most of these adducts are unstable and could be only existed in transition state; third, several competing DNA repair pathways are involved in removing these adducts. To

this end, we chemically synthesized and characterized the DNA glycation adducts arising from Glu-6-P to confirm their identity. Furthermore, their stability was evaluated and proper assays were accommodated to test their biological consequences during DNA replication.

Cellular DNA replication study using M13 vector with site-specific incorporation of N^2-FN6P-dG adducts revealed that the two anomers exhibited strong block (22% and 24% bypass respectively) to polymerase comparing to dG (100% bypass). This result suggests that the two bulky FN6P adducts are poorly tolerated by replication and translesion polymerases, and the two glycation adducts occurring at N2 of Watson-Crick base paring face of dG may impose a large steric hindrance to DNA replication. On the other hand, our mutagenicity test demonstrated that the two N^2-FN6P-dG adducts induced significant G to T and single nucleotide deletion mutations (26% G →T, 30% -1 G for adduct-1, 16% G →T, 36% -1 G for adduct-2) comparing to dG (no mutations detected). These observations indicate the bypass of the two glycation adducts by polymerase is error-prone and in a low fidelity manner. Previously, Yuan et al. have identified methylglyoxal induced N^2-carboxyethyl-dG DNA glycation adducts are weakly mutagenic and only lead to minor G to T mutation (<3%); besides, they have found that the bypass of these adducts are well tolerated by DinB (polymerase IV) in *E. coli* cells[2]. There are a number of previous findings that DinB is capable of bypassing many N^2-dG adducts[13,20–22], we doubted that this polymerase may also be involved in bypassing N^2-FN6P-dG adducts. We performed bypass and mutagenesis studies in Δ*dinB* (pol IV-deficient) *E. coli* cells. However, results obtained from these studies did not show a significant difference of bypass efficiency

compared to wild-type cells (data not shown). Future experiments in testing other translesion polymerases, such as pol II and pol V, may provide insights into the detailed cellular bypass mechanisms on these N^2-FN6P-dG adducts.

Taken together, the results presented in current study provide important evidence showing that N^2-FN6P-dG adducts arising from Glu-6-P, could lead to strong replication block and mutagenicity if left unrepaired, thereby affecting the genome integrity of organisms. In addition, Glu-6-P, cellular fructose 6-phosphate (Fru-6-P, μM level) and ribose 5-phosphate (Rib-5-P, μM level) are also highly reactive glycating agents.[23] The two molecules are 1260 fold for Fru-6-P and 1878 fold for Rib-5-P more powerful than glucose for glycating DNA.[24] The deleterious biological effects generated from these reducing sugar metabolites may represent a unique perspective to study mutagenesis of endogenous molecules.

FIGURES

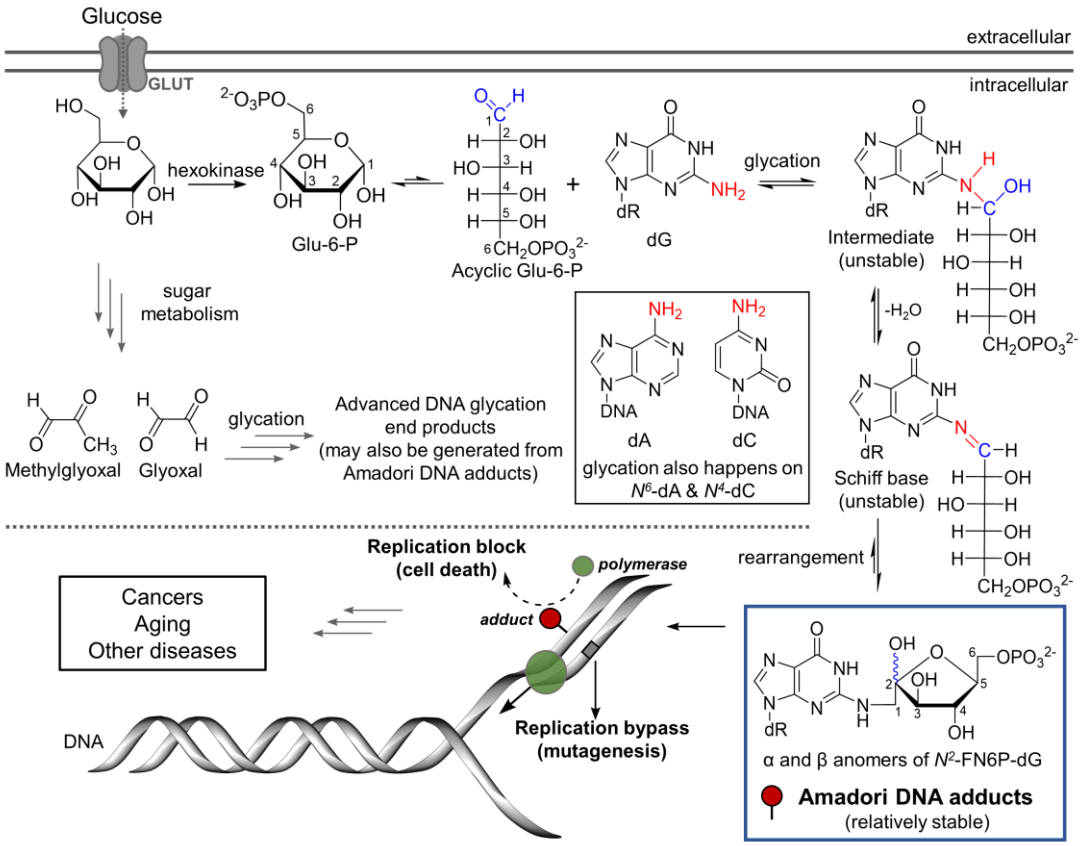

Figure 1. DNA glycation adducts generated from glucose 6-phosphate and their biological implications in cell replication and diseases.

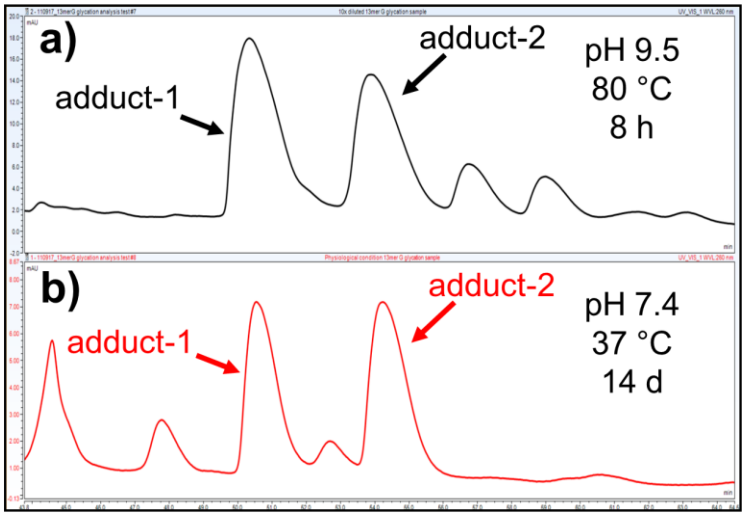

Figure. 2. HPLC separations of glycation oligos generated under **a)** optimized condition and **b)** physiological condition.

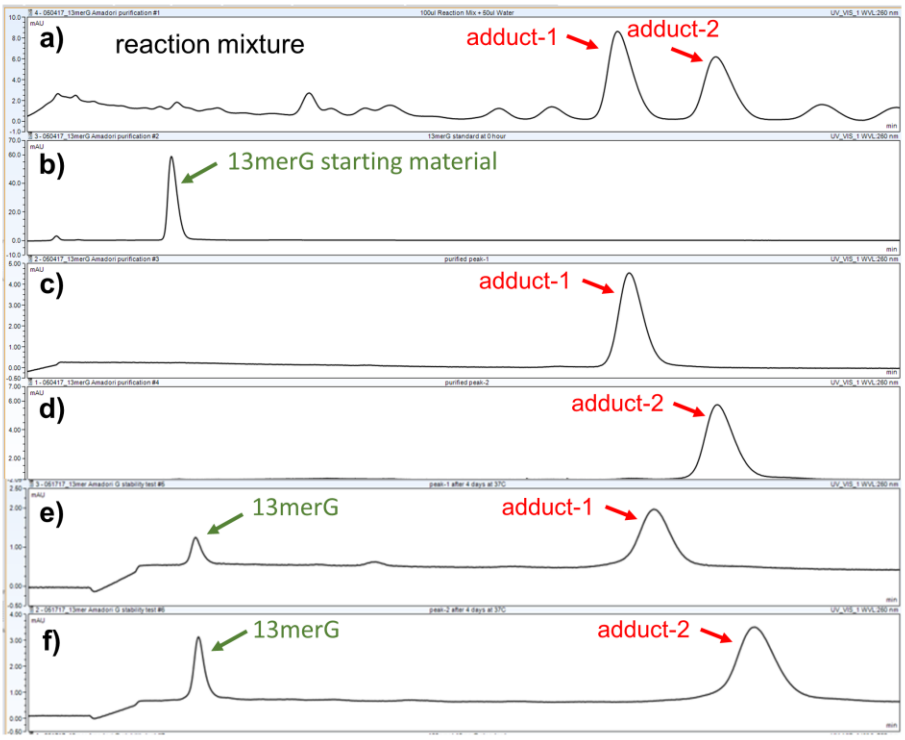

Figure. 3. Separations of DNA oligos by AEX-HPLC. **a)** reaction mixture, **b)** 13mer starting material, **c)** purified adduct-1, **d)** purified adduct-2, **e)** adduct-1 and **f)** purified adduct-2 incubated under 37 ℃ and pH 7.4 for 4 days.

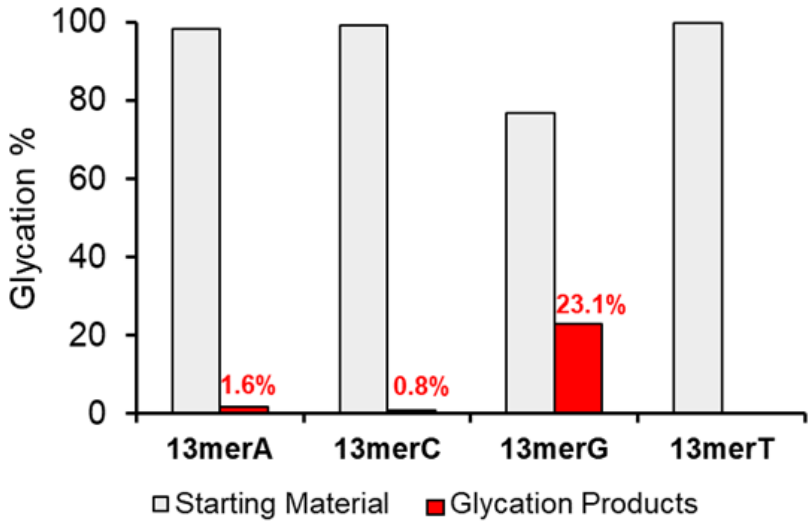

Figure. 4. Glycation adduct formation on four nucleobases.

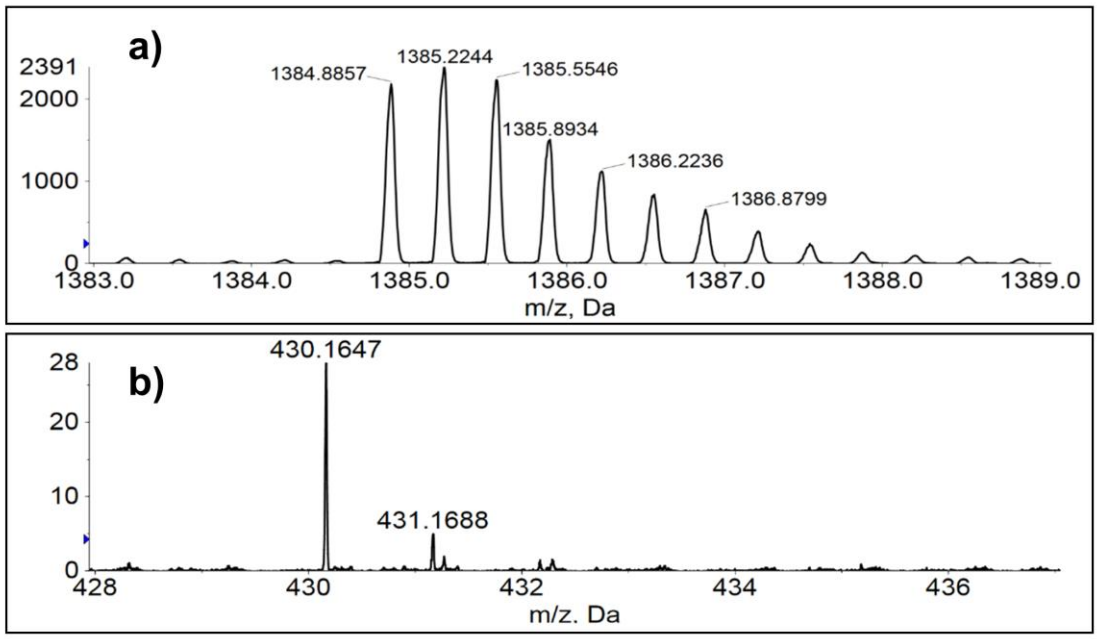

Figure. 5. ESI-TOF analysis of **a)** the N2-FN6P-dG containing 13mer oligo; **b)** N2-FN6P-dG nucleoside product generated from enzymatic digestion and dephosphorylation.

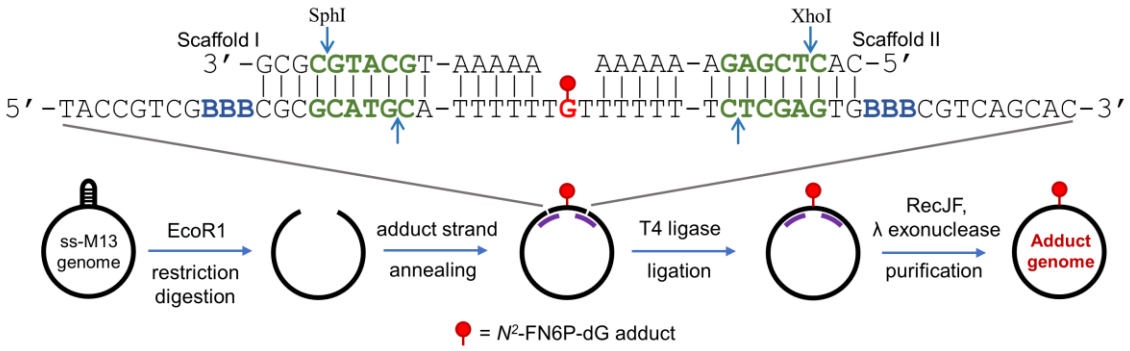

Figure. 6. Genome construction of vectors containing glycation adducts for E. coli assays.

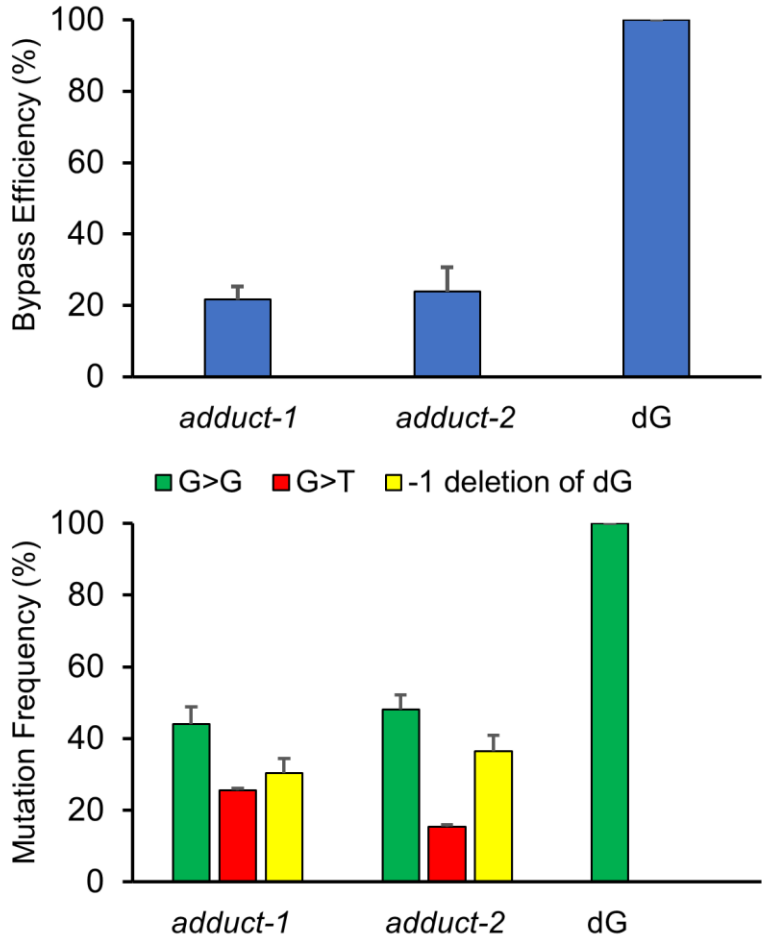

Figure. 7. In cell demonstration of the bypass (top) and mutagenicity (bottom) of the glycation adducts.

ACKNOWLEDGMENTS

The authors want to thank the RI-INBRE program, its director Prof. Bongsup Cho, and staff Dr. Al Bach, Kim Andrews, and Patricia Murray for their kind help. This work was supported by an Institutional Development Award from the National Institute of General Medical Sciences of the National Institutes of Health under grant number P20 GM103430. This work was also supported by National Institutes of Health under grant numbers R15 CA213042 and R01 ES028865 (to D.L.).

REFERENCES

(1) Lee, A. T.; Cerami, A. The Formation of Reactive Intermediate(s) of Glucose 6-Phosphate and Lysine Capable of Rapidly Reacting with DNA. *Mutat. Res.* **1987**, *179* (2), 151–158.

(2) Yuan, B.; Cao, H.; Jiang, Y.; Hong, H.; Wang, Y. Efficient and Accurate Bypass of N^2-(1-Carboxyethyl)-2′-Deoxyguanosine by DinB DNA Polymerase in Vitro and in Vivo. *Proc Natl Acad Sci USA* **2008**, *105* (25), 8679. https://doi.org/10.1073/pnas.0711546105.

(3) Lee, A. T.; Cerami, A. In Vitro and in Vivo Reactions of Nucleic Acids with Reducing Sugars. *Mutat. Res.* **1990**, *238* (3), 185–191.

(4) Yamashiro, T.; Murata, K.; Kawai, S. Extremely High Intracellular Concentration of Glucose-6-Phosphate and NAD(H) in Deinococcus Radiodurans. *Extremophiles* **2017**, *21* (2), 399–407. https://doi.org/10.1007/s00792-016-0913-z.

(5) Pischetsrieder, M.; Seidel, W.; Münch, G.; Schinzel, R. N(2)-(1-Carboxyethyl)Deoxyguanosine, a Nonenzymatic Glycation Adduct of DNA,

Induces Single-Strand Breaks and Increases Mutation Frequencies. *Biochem. Biophys. Res. Commun.* **1999**, *264* (2), 544–549. https://doi.org/10.1006/bbrc.1999.1528.

(6) Veiga da-Cunha, M.; Jacquemin, P.; Delpierre, G.; Godfraind, C.; Théate, I.; Vertommen, D.; Clotman, F.; Lemaigre, F.; Devuyst, O.; Van Schaftingen, E. Increased Protein Glycation in Fructosamine 3-Kinase-Deficient Mice. *Biochem. J.* **2006**, *399* (2), 257–264. https://doi.org/10.1042/BJ20060684.

(7) Fortpied, J.; Maliekal, P.; Vertommen, D.; Van Schaftingen, E. Magnesium-Dependent Phosphatase-1 Is a Protein-Fructosamine-6-Phosphatase Potentially Involved in Glycation Repair. *J. Biol. Chem.* **2006**, *281* (27), 18378–18385. https://doi.org/10.1074/jbc.M513208200.

(8) Van Schaftingen, E.; Delpierre, G.; Collard, F.; Fortpied, J.; Gemayel, R.; Wiame, E.; Veiga-da-Cunha, M. Fructosamine 3-Kinase and Other Enzymes Involved in Protein Deglycation. *Adv. Enzyme Regul.* **2007**, *47*, 261–269. https://doi.org/10.1016/j.advenzreg.2006.12.002.

(9) Wuenschell, G. E.; Tamae, D.; Cercillieux, A.; Yamanaka, R.; Yu, C.; Termini, J. Mutagenic Potential of DNA Glycation: Miscoding by (R)- and (S)-N2-(1-Carboxyethyl)-2'-Deoxyguanosine. *Biochemistry* **2010**, *49* (9), 1814–1821. https://doi.org/10.1021/bi901924b.

(10) Lee, A. T.; Cerami, A. Elevated Glucose 6-Phosphate Levels Are Associated with Plasmid Mutations in Vivo. *Proc. Natl. Acad. Sci. U.S.A.* **1987**, *84* (23), 8311–8314.

(11) Bucala, R.; Model, P.; Russel, M.; Cerami, A. Modification of DNA by Glucose 6-Phosphate Induces DNA Rearrangements in an Escherichia Coli Plasmid. *Proc. Natl. Acad. Sci. U.S.A.* **1985**, *82* (24), 8439–8442.

(12) Dutta, U.; Cohenford, M. A.; Guha, M.; Dain, J. A. In Vitro Nonenzymatic Glycation of DNA Nucleobases: An Evaluation of Advanced Glycation End Products under Alkaline PH. *Anal Bioanal Chem* **2006**, *386* (6), 1633–1640. https://doi.org/10.1007/s00216-006-0753-2.

(13) Delaney, J. C.; Essigmann, J. M. Assays for Determining Lesion Bypass Efficiency and Mutagenicity of Site-Specific DNA Lesions in Vivo. *Meth. Enzymol.* **2006**, *408*, 1–15. https://doi.org/10.1016/S0076-6879(06)08001-3.

(14) Dutta, U.; Cohenford, M. A.; Dain, J. A. Nonenzymatic Glycation of DNA Nucleosides with Reducing Sugars. *Anal. Biochem.* **2005**, *345* (2), 171–180. https://doi.org/10.1016/j.ab.2005.07.034.

(15) John, W. G.; Lamb, E. J. The Maillard or Browning Reaction in Diabetes. *Eye (Lond)* **1993**, *7 (Pt 2)*, 230–237. https://doi.org/10.1038/eye.1993.55.

(16) Moeller, B. C.; Lu, K.; Doyle-Eisele, M.; McDonald, J.; Gigliotti, A.; Swenberg, J. A. Determination of N2-Hydroxymethyl-DG Adducts in the Nasal Epithelium and Bone Marrow of Nonhuman Primates Following 13CD2-Formaldehyde Inhalation Exposure. *Chem. Res. Toxicol.* **2011**, *24* (2), 162–164. https://doi.org/10.1021/tx1004166.

(17) Lu, K.; Ye, W.; Gold, A.; Ball, L. M.; Swenberg, J. A. Formation of S-[1-(N2-Deoxyguanosinyl)Methyl]Glutathione between Glutathione and DNA Induced

by Formaldehyde. *J. Am. Chem. Soc.* **2009**, *131* (10), 3414–3415. https://doi.org/10.1021/ja808048c.

(18) Lu, K.; Craft, S.; Nakamura, J.; Moeller, B. C.; Swenberg, J. A. Use of LC-MS/MS and Stable Isotopes to Differentiate Hydroxymethyl and Methyl DNA Adducts from Formaldehyde and Nitrosodimethylamine. *Chem. Res. Toxicol.* **2012**, *25* (3), 664–675. https://doi.org/10.1021/tx200426b.

(19) Delaney, J. C.; Essigmann, J. M. Mutagenesis, Genotoxicity, and Repair of 1-Methyladenine, 3-Alkylcytosines, 1-Methylguanine, and 3-Methylthymine in AlkB Escherichia Coli. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101* (39), 14051–14056. https://doi.org/10.1073/pnas.0403489101.

(20) Rechkoblit, O.; Zhang, Y.; Guo, D.; Wang, Z.; Amin, S.; Krzeminsky, J.; Louneva, N.; Geacintov, N. E. Trans-Lesion Synthesis Past Bulky Benzo[a]Pyrene Diol Epoxide N2-DG and N6-DA Lesions Catalyzed by DNA Bypass Polymerases. *J. Biol. Chem.* **2002**, *277* (34), 30488–30494. https://doi.org/10.1074/jbc.M201167200.

(21) Zhang, Y.; Wu, X.; Guo, D.; Rechkoblit, O.; Wang, Z. Activities of Human DNA Polymerase Kappa in Response to the Major Benzo[a]Pyrene DNA Adduct: Error-Free Lesion Bypass and Extension Synthesis from Opposite the Lesion. *DNA Repair (Amst.)* **2002**, *1* (7), 559–569.

(22) Jarosz, D. F.; Godoy, V. G.; Delaney, J. C.; Essigmann, J. M.; Walker, G. C. A Single Amino Acid Governs Enhanced Activity of DinB DNA Polymerases on Damaged Templates. *Nature* **2006**, *439* (7073), 225–228. https://doi.org/10.1038/nature04318.

(23) Nakayama, Y.; Kinoshita, A.; Tomita, M. Dynamic Simulation of Red Blood Cell Metabolism and Its Application to the Analysis of a Pathological Condition. *Theor Biol Med Model* **2005**, *2*, 18. https://doi.org/10.1186/1742-4682-2-18.

(24) MORITA, J.; KASHIMURA, N. The Maillard Reaction of DNA with D-Fructose 6-Phosphate. *Agricultural and Biological Chemistry* **1991**, *55* (5), 1359–1366. https://doi.org/10.1271/bbb1961.55.1359.